Derechos y desafíos de la inteligencia artificial

Rights and challenges of Artificial Intelligence

Mariano Enrique Torres Ponce

Abogado y Especialista en Derecho Informático

ORCID: 0009-0009-6284-3867

Diciembre de 2019

Publicado originalmente en *Técnica Administrativa*, 18(1)
Ciencia y Técnica Administrativa · ISSN 1666-1680

Edición navegable y accesible. El contenido se mantiene sin modificaciones.

Resumen

La investigación se centra en la relación de la Inteligencia Artificial con la Humanidad. Brecha: La investigación focaliza particularmente sobre el vacío legal existente a nivel mundial respecto a la Inteligencia Artificial. Propósito: Constituir una base de referencias bibliográficas que asistan a quienes deben elaborar y aplicar las leyes que actúen sobre cuestiones de la Inteligencia Artificial. Metodología: Se ha realizado una revisión de literatura científica de forma sistemática. Resultados: El trabajo expone una vasta y exhaustiva bibliografía sobre lo variado y amplio del desarrollo científico y tecnológico, como así también de sus diferentes problemáticas. Conclusión: Es necesario determinar que los robots deberán tener sus derechos, con foco en su desarrollo y en la interacción con los humanos a fin de atenuar o eliminar los efectos adversos que puedan producirse.

Palabras clave

gobernanza de internet, derecho del ciberespacio, derecho informático, inteligencia artificial, robótica, aprendizaje automático, redes neuronales, agentes inteligentes, ética tecnológica, responsabilidad civil automatizada, privacidad y datos personales, big data, sistemas autónomos, regulación comparada, persona electrónica, convivencia humano-máquina

Abstract

The research focuses on the relationship between Artificial Intelligence and Humanity. Gap: The research focuses particularly on the legal vacuum that exists worldwide regarding Artificial Intelligence. Purpose: To constitute a bibliographical reference base that assists those who must develop and apply laws that act on issues of Artificial Intelligence. Methodology: A systematic review of scientific literature has been carried out. Results: The work presents a vast and exhaustive bibliography on the variety and breadth of scientific and technological development, as well as its different problems. Conclusion: It is necessary to determine that robots should have their rights, with a focus on their development and on the interaction with humans in order to mitigate or eliminate the adverse effects that may occur.

Keywords

internet governance, cyberspace law, computer law, artificial intelligence, robotics, machine learning, neural networks, intelligent agents, technology ethics, automated liability, privacy and personal data, big data, autonomous systems, comparative regulation, electronic personhood, human-robot interaction

Índice

Resumen **2**
Palabras clave 2

Abstract **3**
Keywords 3

1 Introducción **6**
1.1 ¿Qué es natural y qué es artificial? 6

2 Primeros conceptos de inteligencia artificial **8**

3 Tipos de inteligencia artificial en base a su comportamiento **10**
3.1 Sistemas que piensan racionalmente: Sistemas expertos 10
3.2 Sistemas que piensan como humanos: Redes neuronales 13
3.3 Sistemas que actúan racionalmente: Agentes inteligentes 15
3.4 Sistemas que actúan como humanos: Robótica 17

4 Clasificación de inteligencia artificial según su desarrollo **20**
4.1 Inteligencia artificial débil o estrecha 20
4.2 Inteligencia artificial fuerte o general 21
4.3 Súper inteligencia artificial 22

5 Métodos de aprendizaje de la inteligencia artificial **24**
5.1 Machine learning 24
5.2 Deep learning 26

6 La caja negra y el inconveniente de justificar las decisiones **28**

7 Advertencias de un futuro complejo y peligroso **30**

8 Las habilidades humanas de las nuevas máquinas **32**
8.1 Los sistemas aprenderán a mentir 32
8.2 Los robots se harán pasar por humanos 33
8.2.1 Test de Turing 34

8.2.2 Test de Turing en reversa. El surgimiento del CAPTCHA 35
8.3 Los sistemas nos conocerán mejor que otras personas 36
8.4 Las máquinas se parecen físicamente a nosotros 37
8.5 Los sistemas tienen emociones 39
8.6 Las máquinas sustituyen al trabajo humano 40
8.7 La inteligencia artificial se implantará en los cerebros humanos 43
8.8 Los robots como armas militares 44
8.9 Los mismos sistemas serán hackers 46
8.10 La inteligencia artificial con dominio económico 47
8.11 Los sistemas desarrollan su propia ética y moral 48
8.11.1 Las tres leyes de la robótica 50

9 Los veintitrés principios de Asilomar 52

10 La defensa en juicio de los robots 54

11 Marco regulatorio para la inteligencia artificial 56
11.1 Corea del Sur, el precursor de la escena regulatoria 57
11.2 Japón y sus ambiciosos proyectos de legislación ética 59
11.3 China: El gigante desregulado 59
11.4 Unión Europea: Un completo proyecto conjunto 60
11.4.1 Los robots deberán tener un interruptor de emergencia 61
11.4.2 Los robots no podrán hacer daño a los seres humanos 62
11.4.3 No podrán generarse relaciones emocionales con los robots . . 62
11.4.4 Los más grandes deberán tener un seguro obligatorio 64
11.4.5 Derechos y obligaciones para los robots 65
11.4.6 Postergación de las obligaciones tributarias 66
11.5 Alemania y su compromiso con la industria automotor 67
11.6 Estados Unidos: Legalmente indiferente a la ética robot 68

12 ¿Cuál será la mejor personería jurídica para un robot? 70

13 Derechos de humanos para los robots: Sophia, un androide con ciudadanía 73

14 Conclusión 76

Bibliografía 77

Capítulo 1

Introducción

En este trabajo se pretende exponer, de forma sintética y clara, las nuevas situaciones sociales que se proyectan para un futuro cercano y el vacío legal que existe a nivel mundial respecto al creciente fenómeno de la inteligencia artificial. Para esto, intentaremos exhibir lo variado y amplio del desarrollo tecnológico y sus diferentes problemáticas, con la intención de concientizar y, de alguna manera, asistir a aquellos que deben crear las leyes acordes.

1.1 ¿Qué es natural y qué es artificial?

Esta distinción es una de las bases para poder comprender a qué nos estamos refiriendo cuando hablamos de *inteligencia artificial*. En este sentido, lo primero es dilucidar de manera inequívoca qué es *natural* y qué es *artificial*.

El concepto de *natural* podemos verlo adaptado a varios contextos, pudiendo ser adecuado a la forma de actuar, el lugar de nacimiento o pertenencia de una persona o animal, entre otras variantes de su uso. Pero la definición que nos interesa es la más general, la cual nos señalará como *natural* a algo perteneciente o relativo a la naturaleza, o conforme a la cualidad o propiedad de las cosas (Real Academia Española, s. f.). Pudiendo interpretarse como algo fijo o determinado por la naturaleza; relativo a la constitución de una cosa; perteneciente al carácter nativo; de acuerdo a la naturaleza (Periódico de Salud, s. f.).

El inconveniente surge al hablar de *artificial*. El mismo término parecería que nos remite a algo que carece de valor. Donde el propio *Diccionario de la Real Academia Española* define como *artificial* a “algo hecho por mano o arte del hombre”, pero también, en una segunda acepción, lo señala como “no natural, falso” (Real Academia Española, s. f.).

Esto plantea dos cuestiones destacables. La primera es la inevitable asociación que

provoca la definición, dándole un carácter que podría interpretarse como negativo y sumergirnos en un debate interpretativo amplio. En tanto, la segunda cuestión es más amplia y rica en su pensamiento, preguntándonos si realmente "la mano del hombre" al intervenir en la naturaleza vuelve las cosas en *artificiales*. Parece innegable señalar que la construcción de una casa es algo *artificial*, pero, ¿es realmente lógico pensar que el panal que hacen las abejas con celdillas de cera para contener sus larvas y acoplar miel o, por su parte, los nidos que hacen los pájaros donde ponen sus huevos y crían sus pichones, entre otros "animales arquitectos", son diferentes a la mano del hombre?

El ser humano es un animal como cualquier otro interactuando con su medio. Por lo tanto, no vamos a cuestionar, aunque tendríamos herramientas, el concepto de *artificial*, pero sí intentaremos quitarle la visión negativa y considerar a lo *artificial* como parte de lo *natural*.

Los desarrollos del hombre nacen en base a prueba y error, siendo parte de la humanidad como ser pensante. De esta manera, resulta entendible que algunos inventos puedan no hacerle bien a la naturaleza, a diferencia de los animales que instintivamente producen su alimento o crean sus refugios con un mínimo o nulo impacto en el medio ambiente. Pero las creaciones del ser humano son, indudablemente, parte de la misma naturaleza.

Capítulo 2

Primeros conceptos de inteligencia artificial

La *inteligencia artificial* (IA) es una de las ramas de la Informática, con fuertes raíces en otras áreas como la lógica y las ciencias cognitivas. La misma, como primera noción, podemos definirla como un conjunto de técnicas que tienen por objeto dotar a un sistema informático de la capacidad de simular algunas características que se suponen propias de la *inteligencia humana* (Real Academia de Ingeniería, s. f.). Esto podemos fácilmente relacionarlo a aquellos supuestos donde una máquina imita las funciones "cognitivas" que los seres humanos asocian con otras mentes humanas como, por ejemplo: "aprender" y "resolver problemas" (Russell y Norvig, 2009).

Este término fue acuñado en 1956 por John McCarthy, del Instituto de Tecnología de Massachusetts (MIT). En ese año se celebró la conferencia de Dartmouth, en Hanover (Estados Unidos). En ella, McCarthy, Marvin Minsky, Nathaniel Rochester y Claude Shannon establecieron las bases de la *inteligencia artificial* como un campo independiente dentro de la informática. Aunque, previamente en 1950, Alan Turing había publicado un artículo en la revista *Mind*, titulado "Computing Machinery and Intelligence", o de su traducción "Ordenador e inteligencia". En él reflexionaba sobre el concepto de *inteligencia artificial* y establecía lo que luego se conocería como el *test de Turing*, una prueba que permite determinar si una computadora se comporta conforme a lo que se entiende como artificialmente inteligente (Microsoft Corporation, 2009), que después veremos con más detalle.

Desde aquella conferencia en los años 50 hasta la actualidad han pasado varios sucesos, innumerables descubrimientos importantes han marcado un camino, con muchas fallas y aciertos por parte de los desarrolladores encargados de investigar estas ciencias. Pero podemos señalar como la primera gran demostración de la existencia de *inteligencia artificial* para el ciudadano común el 3 de mayo de 1997. En aquel entonces, se produjo la partida de ajedrez más emocionante de la historia. Gary Kasparov,

campeón mundial por más de una década, se enfrentó con *Deep Blue*, la supercomputadora de IBM capaz de calcular 200 millones de posiciones y movimientos por segundo (IBM Corporation, 1997). De los seis juegos, *Deep Blue* ganó tres, Kasparov dos y uno terminó en tablas. Después de muchos años de intentos fallidos, una máquina por fin había sido capaz de doblegar la mente de uno de los hombres más brillantes del siglo. La lejana leyenda de ciencia ficción, en la que el ser humano creaba una máquina capaz de vencerlo, ya era realidad (Hernández Bonilla, 2017).

Después de este evento, el entusiasmo surgió nuevamente sobre las posibilidades que nos brindaban estas tecnologías, otorgándonos mucho más que un partido de ajedrez, sino la perspectiva de crecimiento sin límites en la materia.

Capítulo 3

Tipos de inteligencia artificial en base a su comportamiento

La obra *Inteligencia Artificial: Un Enfoque Moderno*, creada por Stuart Russell y Peter Norvig, es el libro más popular referente a esta materia (Gold, 2011). El mismo, para una mejor comprensión, distingue cuatro tipos de *inteligencia artificial* según sus características (Russell y Norvig, 2009):

- Sistemas que piensan racionalmente. Sistemas expertos.
- Sistemas que piensan como humanos. Redes neuronales.
- Sistemas que actúan racionalmente. Agentes inteligentes.
- Sistemas que actúan como humanos. Robótica.

Donde recorreremos, a lo largo de este capítulo, los detalles del funcionamiento de los diversos sistemas, con sus diferentes ventajas.

3.1 Sistemas que piensan racionalmente: Sistemas expertos

Los *sistemas que piensan racionalmente* son aquellos que buscan emular el pensamiento lógico racional del ser humano para otorgarle la capacidad de percibir, razonar y actuar como estos. La definición se focaliza en el razonamiento, pero partiendo de la premisa de que existe una forma racional de razonar. La lógica permite la formalización del razonamiento y se utiliza para este objetivo.

Los principales exponentes de este tipo de inteligencia son los denominados *sistemas expertos*. Se trata de programas informáticos que permiten aplicar el proceso de

razonamiento de un ser humano experto en una materia para la resolución de problemas específicos con el fin último de tomar decisiones ponderadas en base al conocimiento. Su función es realizar trabajos similares a los que desarrollaría un especialista en un área determinada, sin sustituir a los expertos, sino apoyándolos en un dominio de aplicación específico (Madruga, 2007).

En otras palabras, se trata de un software encargado de capturar el saber de una o varias personas muy experimentadas en una materia e imitar sus procesos de razonamiento cuando soluciona las dificultades de un determinado dominio.

Los *sistemas expertos* tienen la capacidad de resolver problemas complejos de igual o mejor forma que los seres humanos. Esto se debe a que razonan heurísticamente (o sea, simplifican el problema para buscar una solución rápida), usando reglas empíricas efectivas, e interactúan con los humanos de forma adecuada, incluyendo el lenguaje natural, consiguiendo contemplar múltiples hipótesis en competición simultánea y pudiendo justificar sus conclusiones.

Estos sistemas se hallan integrados por tres elementos básicos (Pérez Luño, s. f.):

- *La base de conocimientos*, que funciona como depósito estructurado de toda su información. Esta se desglosa en conocimiento descriptivo (información sobre hechos, situaciones, datos, normas, sentencias, doctrinas, entre otros posibles elementos) y conocimiento procedimental (integrado por reglas de producción heurística del tipo "si... entonces").

- *El motor de inferencia*, que es un software encargado de la correcta aplicación de las reglas de inferencia en las deducciones automáticas. Controla cuándo y cómo se debe aplicar la información contenida en la base de conocimientos para responder a los problemas planteados.

- La *interfaz*, que permite el enlace y la comunicación con el usuario o el ingeniero del conocimiento.

La medida del desarrollo de dicho sistema no es la cantidad de instrucciones o programas, sino la cantidad de reglas contenidas en su base de conocimientos.

El proceso de desarrollo de un sistema experto requiere, primero, la elección del área de interés (por ejemplo, el ámbito del derecho, donde existen los sistemas expertos legales). Una vez hecho esto, se seleccionan los especialistas en la materia (en el caso legal, un juez, un abogado, etc.). Los especialistas, que cuentan con un dominio específico, brindarán conocimientos sobre ese terreno (temas de derecho informático, comercial, laboral, penal, entre otros).

Cuando los expertos aceptan colaborar, el ingeniero de conocimientos extrae su saber y le da una forma de representación adecuada (reglas de producción, redes semánticas o marcos/*frames*), conformando así la base de conocimientos del sistema experto.

SI "premisa" ENTONCES "conclusión" (SI A ENTONCES B)

Cuando las premisas o conclusiones tienen más de un hecho, se conectan con "Y" u "O":

SI *hecho1* Y/O *hecho2* Y/O... *hechoN* ENTONCES *hecho1* Y/O *hecho2* Y/O... *hechoN*

Ejemplos de hechos:

- Juan es un estudiante
- Juan tiene diez años

Una regla combina hechos y permite representar conocimientos e inferir información adicional. Por ejemplo:

R1: Si Juan es estudiante y Juan tiene diez años, entonces Juan estudia en la primaria.

A partir de hechos conocidos es posible deducir nuevos hechos que representan conocimientos derivados. Estos hechos inferidos no requieren ser absolutamente ciertos, pero deben poseer un alto grado de probabilidad para ser considerados válidos. Por ejemplo, Juan podría no estar en la escuela primaria si hubiese avanzado de curso debido a una capacidad intelectual excepcional, pero estadísticamente la inferencia del sistema sigue siendo la más probable.

En términos generales, los sistemas expertos presentan diversas ventajas y desventajas. Entre sus beneficios destacan la preservación del conocimiento especializado, evitando que se pierda con la muerte o retiro del experto humano, y la posibilidad de elevar la calidad del conocimiento de los especialistas al permitirles acceder a información estructurada y actualizada. Asimismo, pueden manejar grandes volúmenes de datos que exceden la capacidad de memoria y consulta de un ser humano, lo que ayuda a prevenir errores derivados de la sobrecarga informativa. Otra ventaja es su alta velocidad de operación en la actualidad.

Entre sus desventajas se encuentran los altos costos y el tiempo considerable necesario para el desarrollo y puesta a punto del sistema. Su tolerancia a fallos es baja, ya que la información se encuentra concentrada y un error puede afectar a todo el funcionamiento. Además, actualizar un sistema experto requiere reprogramación. Por último, la inteligencia artificial aún no ha logrado construir sistemas capaces de resolver problemas de forma general, aplicar sentido común a situaciones complejas o manejar escenarios ambiguos de manera plenamente satisfactoria.

3.2 Sistemas que piensan como humanos: Redes neuronales

Los *sistemas que piensan como humanos* son aquellos que imitan el pensamiento humano con el fin de automatizar actividades vinculadas al proceso de toma de decisiones, la resolución de problemas y el aprendizaje. Su principal exponente son las *redes neuronales artificiales*.

Las redes neuronales artificiales constituyen la rama de la inteligencia artificial más desarrollada y con mayor proyección. Aunque el concepto no es nuevo, durante décadas su avance fue irregular debido a la enorme capacidad de procesamiento que requerían. La aparición de tecnologías con potencia computacional significativamente superior permitió su resurgimiento y demostró su eficacia en múltiples escenarios.

No es sencillo elaborar una definición exacta de las redes neuronales artificiales, pero pueden describirse como una técnica de procesamiento masivo y paralelo de información que emula características esenciales de la estructura neuronal del cerebro biológico (Enciclopedia eumed.net, s. f.).

La clave de su funcionamiento se basa, como todo en tecnología, en algoritmos, entendidos como un conjunto finito de instrucciones o pasos que permiten ejecutar tareas o resolver problemas (Enciclopedia Libre Universal, s. f.).

Su lógica puede comprenderse mediante la comparación entre el procesamiento de información en una computadora y en el cerebro humano. Una computadora es un sistema compuesto por hardware y software. En cuanto a su composición física, el hardware tiene como "elemento de cálculo" más pequeño los transistores, diminutos dispositivos electrónicos que entregan una señal de salida variable en función de la señal de entrada. Los procesadores actuales contienen millones de transistores interconectados de forma determinada al momento de su fabricación para ejecutar diversas operaciones. El software completa este sistema mediante programas que indican al procesador en qué orden ejecutar cada instrucción, lo que permite alcanzar los resultados esperados. En este sentido, aunque no conozcamos el resultado final antes de procesarlo, el camino recorrido está previamente definido, por lo que es posible afirmar que el razonamiento de una persona puede ser replicado por la computadora mediante las órdenes programadas.

Respecto al cerebro, su unidad mínima de procesamiento es la neurona. Se trata de células nerviosas compuestas por un cuerpo celular y diversas prolongaciones, una de las cuales, de forma alargada, es el axón o neurita (Real Academia Española, s. f.). Las neuronas se comunican entre sí mediante impulsos eléctricos que transmiten información. Cada situación activa distintos patrones de conexión, modifica la cantidad de

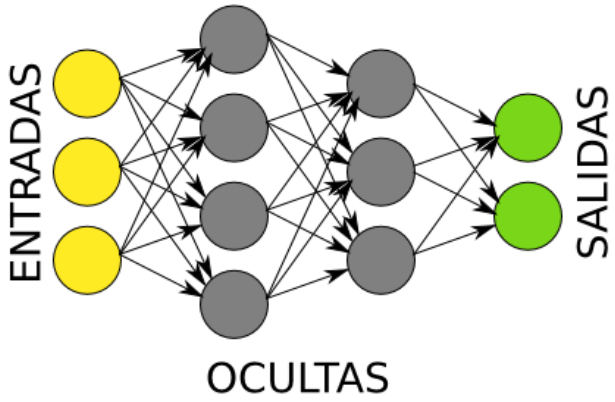
enlaces y altera los flujos de entrada y salida de información según las necesidades del procesamiento.

Al comparar ambas estructuras se advierte que las redes neuronales artificiales buscan imitar la flexibilidad del cerebro. A diferencia de los procesadores tradicionales, estas redes pueden reconfigurar su estructura física y lógica para adaptarse al problema que deben resolver, y pueden diferenciar entre *programación* y *aprendizaje*.

En la actualidad, las redes neuronales artificiales operan sobre grandes infraestructuras de servidores organizadas en centros de datos capaces de manejar enormes volúmenes de información. Los conocimientos incorporados no se limitan a aquello que un programador definió de antemano, sino que se amplían mediante procesos de aprendizaje automático que permiten al sistema superar los límites del conocimiento inicial.

La arquitectura básica de una red neuronal se organiza en tres capas de neuronas:

- La capa de entrada, que recibe la información proveniente del exterior.
- La capa intermedia u oculta, que procesa la información.
- La capa de salida, que entrega el resultado del proceso.



Arquitectura de una red de neuronas

El aprendizaje puede basarse en técnicas de *machine learning* o *deep learning*, conceptos que se desarrollarán más adelante.

Entre las principales ventajas de este tipo de inteligencia artificial se destacan la capacidad de aprender sin necesidad de incorporar cantidades ingentes de datos de manera explícita, la velocidad de procesamiento y su alta tolerancia a fallos. Esta robustez se debe a que el conocimiento está distribuido entre muchas neuronas de la red, lo que permite que el sistema continúe funcionando incluso ante errores o información incompleta.

La principal desventaja es, sin duda, la imposibilidad de explicar la toma de decisiones del sistema. Debido a la complejidad y opacidad de las transformaciones que ocurren en las capas intermedias u ocultas, no es posible determinar cómo llega la red a determinadas conclusiones. Por ello se las describe como una *caja negra*, y quienes trabajan con estas redes plantean que "es un misterio la manera en que llega a la conclusión expresada en la salida" (Brenna, 1994; Orellana, 2015).

Finalmente, es necesario distinguir los sistemas que piensan como humanos de los sistemas que piensan racionalmente, es decir, de los sistemas expertos. La diferencia fundamental es que las redes neuronales pueden *aprender* sin intervención de un especialista, a partir de casos reales provenientes de la experiencia acumulada por muchos expertos (Brenna, 1994). Los sistemas expertos, en cambio, no pueden aprender por sí mismos (Orellana, 2015). Otra diferencia relevante es la mayor tolerancia a fallos de las redes neuronales debido a la distribución de la información, mientras que la principal ventaja de los sistemas expertos sobre las redes neuronales es su capacidad para explicar las decisiones tomadas, algo que las redes neuronales aún no pueden ofrecer.

3.3 Sistemas que actúan racionalmente: Agentes inteligentes

Los *sistemas que actúan racionalmente* a todos aquellos que tienen como finalidad emular de forma racional el comportamiento humano para poder dotar de conductas inteligentes a diversos artefactos.

En este contexto se encuentran los *agentes inteligentes*. Estos sistemas, basados en hardware o software, son capaces de percibir su entorno e interactuar con él mediante la implementación de diferentes actuadores. En esencia, se trata de programas especialmente creados para realizar ciertas tareas de manera autónoma dentro de una red, por encargo de un usuario, con la finalidad de ahorrar tiempo en la vigilancia y en la recolección de información de interés (Equipo editorial de CCM, 2014).

Esta clase de inteligencia artificial puede razonar y aprender a partir de la información que recoge, lo que le permite operar en todo momento, ya sea bajo control humano, supervisado por otro agente o de manera completamente autónoma incluso cuando el usuario se encuentra desconectado. Recolecta datos para finalmente informar o aplicar los parámetros otorgados por el usuario.

Los agentes inteligentes se caracterizan por cuatro cualidades fundamentales (Wooldridge & Jennings, 1995):

- Autonomía. Operan sin intervención directa de seres humanos y mantienen con-

trol sobre sus acciones y su estado interno.

- Sociabilidad. Pueden interactuar con otros agentes o con seres humanos mediante un lenguaje o sistema de comunicación.
- Reactividad. Perciben su entorno y responden de forma oportuna a los cambios que se producen en él, ya sea por variaciones en el mundo físico, la intervención de un usuario, la acción de otros agentes, el acceso a Internet o la combinación de estos elementos.
- Proactividad. No solo reaccionan a su entorno, sino que pueden actuar con iniciativa propia, guiados por objetivos establecidos.

El comportamiento de un agente puede representarse mediante una tabla que relaciona la secuencia de percepciones con las acciones que deben ejecutarse en respuesta a cada estímulo, tomando como referencia su historial de conocimientos. Estos agentes se desplazan dentro de un entorno *virtual* y operan mediante un sistema informático, desempeñando una función análoga a la de los robots en el mundo físico. Por ello, en ocasiones se los denomina *softbots* o *software robots*.

Respecto al control del entorno, un agente puede actuar en contextos estáticos o dinámicos. En los entornos variables el control del agente solo es parcial, por lo que una misma acción puede producir resultados diferentes en ocasiones distintas.

Los agentes inteligentes se diferencian de los sistemas expertos previamente analizados, dado que los sistemas expertos no suelen interactuar de manera directa con su entorno. Por lo general están diseñados para tareas más complejas dentro de un dominio específico en el que controlan completamente las decisiones y no cooperan entre sí.

Además de facilitar el trabajo de los usuarios, los agentes inteligentes pueden funcionar como consultores y operadores dentro de entornos complejos. Sin embargo, presentan desafíos importantes, como la dificultad para identificar qué información es relevante, la necesidad de optimizar mecanismos de búsqueda, la complejidad de evitar tareas repetidas y la adaptación a múltiples protocolos, formatos y sistemas de acceso a datos.

En términos generales, podemos decir que antiguamente los agentes inteligentes eran muy utilizados. Su funcionamiento se podía observar en motores de búsqueda web para la recuperación de información de Internet o en servicios de soporte en línea y chat. Actualmente, su uso es mucho menor, ya que los desarrollos de inteligencia artificial tienden a concentrar las diferentes soluciones utilizando la tecnología de redes neuronales y los diversos modelos de aprendizaje, provocando un desplazamiento gradual en los distintos rubros de este tipo de sistemas.

3.4 Sistemas que actúan como humanos: Robótica

Los *sistemas que actúan como humanos* pueden entenderse como aquellos que buscan desenvolverse en un contexto similar al de las personas físicas. Imitan su comportamiento con el objetivo de realizar tareas que hoy desempeñan los seres humanos, pero no cualquier actividad, sino aquellas que requieren algún grado de inteligencia. El ejemplo más representativo de esta categoría es la *robótica*.

La robótica puede definirse como la técnica que aplica la informática al diseño y al uso de aparatos que, en sustitución de las personas, ejecutan operaciones o trabajos, principalmente en instalaciones industriales (Real Academia Española, s. f.).

La base de la robótica son los robots. El término *robot* surgió en la década de 1920 en la obra de teatro de ciencia ficción *R. U. R. (Robots Universales Rossum)*, escrita por el checo Karel Čapek. El vocablo proviene de la palabra checa *robota*, que significa trabajo, y expresa una concepción social y tecnológica muy diferente a la moderna (Robotsia, 2015b).

Un *robot* puede describirse como una máquina o ingenio electrónico programable capaz de manipular objetos y ejecutar operaciones que antaño solo realizaban los seres humanos. En este sentido, los robots son creaciones destinadas a asistir a las personas en trabajos difíciles o pesados. A partir de esta noción surgen conceptos relacionados como *cobots*, *androides* y *cyborgs*.

Los *cobots* son robots colaborativos diseñados para interactuar físicamente con los seres humanos. Se emplean en tareas repetitivas que no requieren creatividad, como el manipulado de materiales o el empaquetado de productos. Se caracterizan por ser pequeños, ligeros y de movimientos lentos, y operan bajo una regla fundamental: si un ser humano se acerca demasiado, deben detener su funcionamiento (Vidal, 2017).

El término *androide* alude a un autómata con figura humana. Se trata de un tipo de robot antropomorfo que reproduce aspectos físicos y conductuales de las personas. En un sentido más amplio puede extenderse también a robots zoomórficos y a sistemas de locomoción inspirados en seres vivos.

La última distinción es la de *cyborg*. El concepto fue introducido por los científicos Manfred Clynes y Nathan Kline, quienes lo imaginaron como una combinación de humano y máquina capaz de adaptarse a entornos extraterrestres (Clynes y Kline, 1960). Con el tiempo, la idea evolucionó hacia la noción de un ser humano cuyas funciones corporales han sido modificadas mediante dispositivos o sustancias que le permiten subsistir en ambientes no habituales y mejorar sus capacidades naturales.

Una vez realizadas estas distinciones podemos retomar el objeto de estudio. La robótica, en sus diferentes formas, constituye uno de los principales medios a través de

los cuales la inteligencia artificial interactúa con el entorno, ya adopte un aspecto mecánico, una apariencia humana o un rol asistencial permanente.

Dado que en la actualidad el término robot es ambiguo, es necesario aclarar que aquí se hace referencia a un sistema electromecánico conducido generalmente por un programa informático o por un circuito eléctrico. En cambio, la palabra *bot* se utiliza para describir programas informáticos que realizan tareas repetitivas de manera automática, imposibles o muy tediosas para una persona. A pesar de su nombre, los bots no forman parte del campo de la robótica, ya que aluden a sistemas virtuales basados en software, mientras que los robots implican un componente físico además del lógico (Biobloguer, 2011).

Existe una amplia variedad de robots, pero no todos incorporan inteligencia artificial. Para distinguir cuáles son relevantes desde el punto de vista tecnológico y jurídico, se adopta la clasificación propuesta por la Asociación de Robots Japonesa (JARA), que agrupa los robots en seis categorías según su nivel de inteligencia (Universidad Konan, s. f.; Facultat d'Informàtica de Barcelona, s. f.):

- Dispositivos de manejo manual. Son controlados directamente por una persona.
- Robots de secuencia arreglada. Ejecutan una secuencia preprogramada que permanece invariable hasta su total reprogramación.
- Robots de secuencia variable. Su secuencia de operaciones puede modificarse fácilmente por un operador.
- Robots regeneradores. El operador humano guía al robot durante la realización de la tarea.
- Robots de control numérico. El operador introduce la programación del movimiento hasta que la tarea queda enseñada manualmente.
- Robots inteligentes. Son capaces de comprender e interactuar de manera adaptativa, modificando su comportamiento según los cambios del entorno.

De esta clasificación se consideran pertinentes para este estudio únicamente los denominados *robots inteligentes*.

En el ámbito jurídico, Javier Antonio Nisa Ávila, especialista en Derecho de Tecnologías Avanzadas e Inteligencia Artificial, propone una clasificación complementaria de cuatro niveles de robots inteligentes (Nisa Ávila, 2016). Al tratarse de una aproximación construida desde el derecho, resulta especialmente útil para ordenar y proyectar la regulación existente y futura. Sus niveles son los siguientes:

- Nivel 1 - Sistemas inteligentes programados: Definidos en Japón en la *Robot Strategy* y en la Unión Europea en el proyecto *Regulating Robotics: A Challenge for Europe*, como máquinas que incorporan sistemas robóticos que ayudan a sus operadores o propietarios a realizar tareas automatizadas, sin constituir propiamente robots. Un ejemplo son los vehículos de conducción autónoma.

- Nivel 2 - Robots no autónomos: Según la legislación coreana sobre desarrollo y distribución de robots inteligentes, son dispositivos que realizan tareas programadas simples sin asistencia humana y que pueden tomar decisiones mecánicas dentro del marco de su función. Ejemplos habituales incluyen la robótica sanitaria, la asistencia en el hogar y la limpieza automática.

- Nivel 3 - Robots autónomos: Según la ley coreana y la estrategia japonesa, tienen capacidad para ejecutar tareas complejas, priorizar acciones y tomar decisiones propias dentro del ámbito de trabajo asignado, basadas en objetivos establecidos por un operador. Incluyen sistemas de navegación aérea automatizada, ferroviaria y de diseño industrial autónomo.

- Nivel 4 - Inteligencia artificial: Son sistemas capaces de percibir su entorno sin necesidad de órdenes externas preprogramadas, discernir distintas circunstancias y moverse de forma voluntaria. Pueden obedecer tanto órdenes humanas como señales provenientes de otros sistemas inteligentes.

Esta diferenciación facilita la identificación de las características esenciales de los distintos tipos de robots. No obstante, el uso del término *inteligencia artificial* exclusivamente para el nivel 4 resulta limitado, dado que el concepto abarca un espectro mucho más amplio de sistemas, según se analizará más adelante.

Por último, en la actualidad los robots comerciales e industriales se utilizan de manera extendida y realizan tareas con mayor precisión y menor coste que los humanos. Se los emplea también en actividades peligrosas, insalubres o extenuantes. Sus aplicaciones incluyen la manufactura, el montaje y embalaje, el transporte, la exploración terrestre y espacial, la cirugía, el armamento, la investigación científica y la producción en masa de bienes industriales y de consumo (Macchiavello, 2008).

Otras aplicaciones destacadas abarcan la limpieza de residuos tóxicos, la minería, la búsqueda y el rescate de personas y la localización de minas terrestres. Entre los usos más relevantes se encuentran los vinculados con la medicina y el ámbito militar, sobresaliendo la asistencia a personas con discapacidades y a personas mayores.

Capítulo 4

Clasificación de inteligencia artificial según su desarrollo

La inteligencia artificial puede clasificarse en tres tipos, según su nivel de desarrollo, donde se incrementarán las capacidades del sistema variando en cada una de ellas. Las etapas son: inteligencia artificial débil o estrecha, inteligencia artificial fuerte o general y súper inteligencia artificial. Las cuales serán estudiadas tomando como punto de referencia las tecnologías que actualmente presentan una mayor evolución. Por este motivo, nos basamos en las redes neuronales, para demostrar su desarrollo. Conceptos que iremos explicando de forma detallada a lo largo del presente capítulo.

4.1 Inteligencia artificial débil o estrecha

La expresión inglesa *Artificial Narrow Intelligence* (ANI) designa la inteligencia artificial débil o estrecha. Se trata de sistemas que, mediante redes neuronales de tamaño reducido, buscan resolver problemas concretos sin aspirar a reproducir la inteligencia humana en sentido amplio. Por ello abarcan un campo reducido y especializado.

Su funcionamiento se basa en el aprendizaje. Estos sistemas han sido optimizados mediante técnicas de *machine learning* o *deep learning*, lo que les permite instruirse de forma autónoma sin que sea necesario instalarles un software con una base de datos predeterminada que indique explícitamente sus criterios de comportamiento. Gracias a este proceso, pueden desarrollar la capacidad de ejecutar una o varias tareas específicas a partir de algunas instrucciones iniciales y de un sistema de entrenamiento que gradualmente mejora su desempeño, hasta el punto de superar al ser humano en esa tarea concreta.

Una vez entrenadas, estas redes neuronales pueden combinarse con distintos programas y potenciar su funcionamiento, como ha ocurrido en numerosos casos exitosos.

La inteligencia artificial débil se caracteriza por estar orientada a realizar un conjunto limitado de actividades definidas, sin importar cuán bien las ejecute. *Deep Blue* pertenece a esta categoría: por más que haya logrado un desempeño extraordinario en ajedrez, esa es la única actividad que está capacitado para realizar.

Entre los ejemplos más destacados se encuentra el motor de búsqueda de Google. Durante años empleó un algoritmo matemático complejo que simulaba inteligencia, pero que no dejaba de ser un cálculo estático. Desde 2015 se encuentra potenciado por una inteligencia artificial de este tipo denominada *RankBrain* (Llorca, 2017).

En el ámbito jurídico existe el sistema *ROSS*, desarrollado por IBM y utilizado en una firma estadounidense especializada en la gestión de quiebras. Su función consiste en formular hipótesis que respondan a una pregunta concreta mediante el análisis de todo el cuerpo normativo relevante (Zahumenszky, 2016).

En el campo penal opera el sistema *COMPAS*, que asigna un puntaje a los imputados con el fin de predecir la probabilidad de reincidencia. Se emplea en diversos Estados de los Estados Unidos que utilizan sistemas de evaluación para definir la eventual concesión de la libertad condicional. Además, existen numerosos programas de análisis de mercado para inversiones bursátiles capaces de realizar transacciones en cuestión de segundos, sistemas de conducción autónoma y múltiples aplicaciones similares (Corvalán, 2017).

No obstante, no todos los casos han sido exitosos. Un ejemplo paradigmático es *Tay*, la inteligencia artificial desarrollada por Microsoft en 2016. Se trataba de un chatbot diseñado para conversar con los usuarios y aprender de dichos intercambios con el fin de desarrollar una personalidad propia. Sin embargo, al día siguiente de su lanzamiento, los diálogos mantenidos en redes sociales lo llevaron a reproducir expresiones de carácter nazi y misógino debido a la manipulación ejercida por ciertos usuarios. El resultado ilustra el conocido principio *Garbage In, Garbage Out*: si ingresa basura, saldrá basura. El experimento fracasó y dejó en evidencia que, en casos como este, el aprendizaje requiere la supervisión de personas responsables para incorporar parámetros éticos que distingan entre el bien y el mal (Suárez, 2016).

4.2 Inteligencia artificial fuerte o general

La expresión inglesa *Artificial General Intelligence* (AGI) designa la inteligencia artificial fuerte o general. Esta corriente parte de la idea de que la mente humana está constituida por algoritmos altamente complejos que podrían ser descifrados y convertidos en programas informáticos, lo que permitiría digitalizar y almacenar la mente en un sistema computacional (Madruga, 2015). En este escenario ya no se trataría de una

inteligencia artificial diseñada para resolver problemas puntuales con gran eficiencia, sino de un sistema capaz de abordar problemas generales, tal como indica su denominación.

En un primer momento, una inteligencia artificial fuerte se asemejaría a la mente de un niño pequeño, con capacidad para responder preguntas simples sobre múltiples temas. La cuestión central consiste en determinar si podría llegar a desarrollar *raciocinio*. Este punto abre dificultades significativas, ya que conduce al segundo estadio: el desarrollo.

Una vez en funcionamiento, un sistema de este tipo podría ofrecernos soluciones a problemas cada vez más complejos. Sin embargo, su evolución es incierta, porque resulta imposible conocer con precisión qué procesos ejecuta internamente para producir cada resultado y qué tipo de recableado aleatorio realiza su red neuronal durante la operación. Esta incertidumbre vuelve a poner de relieve el problema de la denominada caja negra.

En este punto puede afirmarse que su comportamiento comenzaría a asemejarse a un razonamiento lógico. Esta similitud permitiría sostener que se está ante una inteligencia artificial fuerte, diferenciándola de la inteligencia artificial débil, cuya capacidad se limita a alcanzar valores predeterminados. Los seres humanos, en cambio, no solo alcanzan esos valores, sino que además son capaces de desarrollarlos y transformarlos (Perissé, 2001). Si bien este tipo de inteligencia todavía no existe, se considera probable que en los próximos años pueda surgir mediante la combinación de tres líneas de investigación: la recreación del funcionamiento de un cerebro, la simulación de la evolución celular y la programación de inteligencias artificiales débiles con capacidad para integrarse y generar un sistema general.

4.3 Súper inteligencia artificial

La *súper inteligencia artificial* constituye el grado más alto de desarrollo imaginable. Puede definirse como un tipo de inteligencia artificial que ya no busca imitar el pensamiento humano, sino superarlo, alcanzando una capacidad cognitiva mayor que la de cualquier persona (Technopedia, s. f.).

El término proviene de la expresión inglesa *Artificial Superintelligence* (ASI). Esta forma de inteligencia emergería como una evolución de la *Artificial General Intelligence* (AGI). Al igual que esta, comenzaría con capacidades comparables a las de un niño pequeño, pero su desarrollo sería exponencial. Su característica más determinante radicaría en la posibilidad de recablear de manera ilimitada su red neuronal para adaptarla a los resultados obtenidos en sus experiencias y experimentos. A ello se sumaría

una capacidad de almacenamiento prácticamente inagotable.

La aparición de un sistema de estas características daría lugar a lo que los científicos denominan *singularidad tecnológica*. Este fenómeno describe una inteligencia artificial que, siendo inicialmente equivalente a un ser humano, pronto lo superaría y adquiriría la capacidad de mejorarse a sí misma, además de diseñar nuevas computadoras igual o más potentes que ella misma. De este modo se generaría un crecimiento de inteligencia fuera de control, capaz de superar indefinidamente los niveles cognitivos humanos.

En un artículo influyente, el filósofo David Chalmers sostiene que la llegada de este fenómeno no es una cuestión hipotética, sino una cuestión de tiempo. Propone una lógica progresiva para describir el ascenso hacia la súper inteligencia, diferenciada en tres niveles (Chalmers, 2010):

- IA: un tipo de inteligencia artificial al menos igual a la humana.
- IA+: inteligencia artificial superior a la humana.
- IA++: inteligencia artificial muy superior a la humana, es decir, súper inteligencia. Según Chalmers, una IA con capacidad de desarrollo podría evolucionar hacia IA+, y una vez alcanzado ese punto, el tránsito hacia IA++ sería inevitable.

A lo largo de los últimos años han surgido confusiones mediáticas acerca del verdadero estado de la inteligencia artificial. En ocasiones se ha presentado a supercomputadoras con una enorme capacidad de procesamiento y almacenamiento como si fuesen formas de inteligencia avanzada, cuando en realidad no poseen la capacidad de adquirir conocimientos nuevos ni de aprender de la experiencia, sino que operan sobre bases de datos estáticas.

Buena parte de este desconcierto se explica por la denominada *paradoja de Moravec*. El investigador austriaco Hans Moravec planteó que resulta relativamente sencillo conseguir que los sistemas exhiban rendimientos similares a los de un adulto en pruebas de inteligencia o en juegos como las damas, pero es extremadamente difícil dotarlos de habilidades perceptivas y motoras comparables a las de un bebé de un año. En otras palabras, las tareas que los humanos perciben como complejas suelen requerir menos desarrollo tecnológico que las habilidades básicas sensoriomotoras que realizamos de forma casi inconsciente (Parra, 2019).

En la actualidad contamos con múltiples inteligencias artificiales débiles extraordinariamente potentes, pero aún queda un largo camino para alcanzar una inteligencia artificial fuerte y será difícil identificar el momento preciso de su aparición. Sin embargo, diversos estudios coinciden en que, una vez lograda una AGI funcional, su progresión hacia la súper inteligencia podría producirse con notable rapidez.

Capítulo 5

Métodos de aprendizaje de la inteligencia artificial

Actualmente existen dos métodos que son predominantes en el área del aprendizaje de estas tecnologías. El primero es *machine learning* o aprendizaje automático, nacido a fines de la década del '70 y principios de los '80. En tanto, en el año 2006, surgiría como rama de aquella el *deep learning* o aprendizaje profundo, acuñado así por Geoffrey Hinton, como forma de explicar nuevas arquitecturas de redes neuronales profundas que son capaces de adquirir los conocimientos mucho mejor empleando modelos más planos. Estos métodos iremos viéndolos en detalle a lo largo del capítulo.

5.1 Machine learning

El aprendizaje automático, conocido en inglés como *machine learning*, es la rama de la inteligencia artificial dedicada al estudio de agentes o programas que aprenden o evolucionan en función de su experiencia con el fin de mejorar progresivamente su desempeño en una tarea específica. Su objetivo principal, como en todo proceso de aprendizaje, consiste en utilizar la evidencia conocida para formular una hipótesis y producir una respuesta adecuada ante situaciones nuevas no previstas (Mitchell, 1997).

Para comprender su funcionamiento es útil contrastar la forma en que los seres humanos y los programas informáticos resuelven problemas. Las personas aprenden inicialmente a partir de la teoría y consolidan ese conocimiento mediante su aplicación práctica. Los programas informáticos, en cambio, requieren la intervención del programador, quien identifica un problema y diseña un sistema que intente abarcar todas las posibles resoluciones. Posteriormente formula instrucciones secuenciales para que la computadora las ejecute, proceso conocido como *parseo de datos*.

El *machine learning* combina ambas capacidades: por un lado, la flexibilidad humana para aprender de nuevas situaciones y, por otro, la precisión del software basado

en algoritmos diseñados para resolver eficientemente casos conocidos. Puede visualizarse su funcionamiento como una caja compuesta por dos bloques. El primero contiene múltiples algoritmos que representan la teoría programada por el desarrollador, que cubre las soluciones conocidas para los problemas más frecuentes. El segundo corresponde al algoritmo de aprendizaje automático, encargado de determinar cómo y cuándo deben emplearse esos algoritmos frente a problemas nuevos.

Para ilustrar este proceso resulta útil compararlo con la educación infantil. El sistema recibe un estímulo o ejemplo, equivalente a cuando un docente explica a un niño la lógica necesaria para resolver un problema. El sistema procesa ese estímulo mediante el algoritmo de aprendizaje y genera un modelo, que equivale a la interpretación del conocimiento teórico. Luego comienza la etapa de entrenamiento, en la que se le proporcionan nuevos casos para que los resuelva. Según acierte o falle, se aplican instrucciones de ajuste al modelo, de forma análoga al método de premio y castigo utilizado en la educación humana. Se trata de un aprendizaje supervisado, ya que requiere ejemplos proporcionados por humanos en la fase inicial (Mitchell, 1997).

Los seres humanos sabemos reconocer si un documento es una queja, una instancia o una solicitud. Una vez que los algoritmos reciben un conjunto de entrenamiento elaborado por personas, pueden generalizar ese conocimiento y clasificar documentos automáticamente sin intervención humana (Rodríguez García, 2017b).

En la práctica, este tipo de aprendizaje se observa en los sistemas que clasifican correos electrónicos y filtran el spam. El ejemplo más destacado es el utilizado por Gmail. Una persona puede clasificar correos casi de forma inconsciente y con un margen mínimo de error. Para que un sistema pueda hacerlo, el programador identifica patrones frecuentes en correos no deseados, como palabras clave, signos de exclamación o errores gramaticales. Aunque estas características no son universales, proporcionan información inicial para que el algoritmo analice el asunto y el cuerpo del correo y logre un alto grado de efectividad.

Con la incorporación de nuevos datos, el sistema continúa perfeccionándose. Reconoce nuevos patrones de spam, identifica variaciones y adquiere conductas que el programador no detectó inicialmente. Paralelamente, también aprende de las correcciones realizadas por los usuarios. Si se reclasifica un mensaje, el sistema ajusta sus prioridades, depura el error y mejora la secuencia de algoritmos de filtrado (Arrabales, 2016).

Otras aplicaciones del aprendizaje automático incluyen la detección de rostros en cámaras digitales y teléfonos móviles, la personalización de resultados en buscadores de Internet, los antivirus con detección de software malicioso, sistemas de recomendación basados en datos de los usuarios y herramientas de predicción o pronóstico (Valencia, 2017).

Actualmente existen numerosos servicios basados en esta tecnología, como Amazon Machine Learning, IBM Watson Developer Cloud y Microsoft Azure Machine Learning, entre otros.

5.2 Deep learning

El aprendizaje profundo, conocido en inglés como *deep learning*, es una derivación dentro del campo del *machine learning* basada en la idea del aprendizaje exclusivamente desde el entrenamiento. En lugar de que el programador proporcione una lista extensa de reglas para resolver un problema, se entrega directamente un modelo capaz de evaluar ejemplos y ajustarse en función de los errores. Con el tiempo se espera que estos modelos logren resolver el problema con un grado de precisión extremadamente elevado, dado que el sistema es capaz de extraer patrones o anomalías por sí mismo (Rodríguez, 2017). Por esta razón se lo considera una forma de aprendizaje no supervisado.

Su funcionamiento puede comprenderse mediante el ejemplo educativo del niño pequeño. En este escenario no existe teoría previa, ni lo que sería para la máquina un parseo de datos. El niño aprende primero la palabra "perro" y luego intenta aplicarla señalando distintos objetos. El adulto confirma o corrige esas respuestas, y con cada interacción el niño se vuelve más consciente de las características comunes de los perros. Construye una abstracción compleja del concepto, formada por niveles que se apoyan en el conocimiento adquirido en los niveles anteriores. El sistema de *deep learning* opera del mismo modo: cada resultado incorpora nuevas características del objeto y mejora progresivamente la precisión del modelo.

El funcionamiento de *deep learning* en una red neuronal se produce mediante el uso de sus capas internas, a menudo denominadas caja negra. Cada capa procesa información y produce un resultado expresado en términos de ponderación. Si una unidad analiza una imagen para identificar un perro, podría estimar que hay un 78 por ciento de probabilidad de que lo sea. La siguiente capa combinará ese dato con su propio análisis, generando una nueva ponderación que se transmitirá a la capa posterior, y así sucesivamente en múltiples niveles. Algunos algoritmos emplean más de cien capas (Bejerano, 2017).

El término "profundo" alude justamente a la cantidad de capas ocultas de la red neuronal. Mientras que las redes tradicionales contienen dos o tres capas, las redes profundas pueden superar ampliamente el centenar (MathWorks, s. f.).

Esta estructura representa una ventaja sobre el *machine learning* tradicional. En el aprendizaje supervisado, el programador debe definir con precisión los atributos ne-

cesarios para que el sistema reconozca un objeto, proceso denominado extracción de características. La eficacia depende de la exactitud de esa definición. En cambio, en el aprendizaje profundo el conjunto de características es construido de manera autónoma por el propio modelo. Esto no solo acelera el proceso, sino que generalmente resulta más preciso (Valencia, 2017).

Otra característica es que el *deep learning* imita el funcionamiento de un sistema nervioso. Cada capa actúa de manera distinta y se especializa en un atributo determinado. A medida que el entrenamiento avanza comienzan a emerger patrones vinculados a funciones humanas como la memoria, el razonamiento, la atención, la motivación y las emociones (Rodríguez García, 2017b).

Esto contrasta con el aprendizaje dirigido, en el cual todas las unidades de la red neuronal se comportan de manera homogénea y pueden desempeñar las mismas funciones.

El aprendizaje profundo es, en consecuencia, la tecnología más empleada y en mayor expansión. Actualmente se utiliza en traductores automáticos, sistemas avanzados de reconocimiento de voz, interpretación semántica en asistentes de voz como Apple Siri y Microsoft Cortana, reconocimiento automático de imágenes y múltiples aplicaciones que dependen del procesamiento complejo de datos.

Capítulo 6

La caja negra y el inconveniente de justificar las decisiones

Como fue adelantado en el capítulo de redes neuronales, el gran problema de este tipo de inteligencia artificial y del desarrollo de técnicas de aprendizaje, principalmente de deep learning, es el fenómeno de la caja negra. Esta contingencia nace de la imposibilidad de conocer qué sucede en el interior de las capas intermedias de las redes neuronales que nos permitiría justificar el resultado o decisión tomada. Es así que, en este contexto solo conocemos qué ocurre en la entrada y en la salida del mismo. Dichas capas son tan opacas como el cerebro. Este impedimento se presenta, debido a que, en lugar de almacenar lo que ha aprendido en un bloque de memoria digital, propaga la información de una forma extremadamente compleja de descifrar.

La caja negra, es el gran desencadenante de los miedos y las incertidumbres sobre el control y el desarrollo que debe darse a estas tecnologías.

El profesor de la Universidad de Carnegie Mellon, en Pensilvania, Dean Pomerleau fue uno de los primeros en enfrentarse a esta dificultad en 1991. En aquel entonces, se encontraba desarrollando un vehículo militar de conducción autónoma.

Luego de varias sesiones de aprendizaje de conducción exitosas, en un recorrido de pruebas, el automóvil se acercó a un puente y se empezó a desviar peligrosamente. Pomerleau, con muy buenos reflejos, tomó el control manual y evitó el impacto. Después del evento, el científico decidió comprender qué ocurrió con esa red neuronal y empezó a estudiar la caja negra, o sea las capas intermedias. Realizando incansables estudios sobre la interpretación de la inteligencia artificial de los diversos estímulos visuales, pudo encontrar qué le hizo creer en aquel momento que la mejor opción era salirse del camino. El sistema tomaba como referencia para determinar el camino, los bordes de pasto que estaban al costado y, en aquella ocasión, el aspecto del puente le resultó confuso.

Pero si, en ese entonces, era complejo interpretar las capas intermedias de una red

neuronal, actualmente podemos decir que se acerca a lo imposible. Dado que los sistemas no son como antes donde predominaba el machine learning, con redes neuronales compuestas con algunas capas y con las indicaciones claras del resultado esperado brindadas por un programador. En estos tiempos la evolución del propio aprendizaje automático y la aparición del aprendizaje profundo, gestionan toda la información de forma desconocida para nosotros y son los algoritmos que se encuentran en la gran cantidad de capas intermedias los que le "autoenseñan" los conceptos adquiridos a esta inteligencia artificial.

Capítulo 7

Advertencias de un futuro complejo y peligroso

El temor a lo desconocido es una característica inherente no solo a la especie humana, sino a todo ser vivo. La incertidumbre respecto del resultado de una acción es un factor natural. Sin embargo, cuando quienes plantean escenarios alarmantes son los propios especialistas, personas que comprenden en profundidad qué implica hablar de *inteligencia artificial*, la preocupación adquiere otra dimensión. Resulta necesario prestar especial atención a figuras reconocidas de la ciencia y la tecnología que han manifestado su descontento y han advertido que la inteligencia artificial debe desarrollarse y administrarse con mayor responsabilidad.

Entre quienes alertan sobre un avance desmedido se encuentra Bill Gates. El cofundador de Microsoft ha expresado en reiteradas ocasiones su preocupación por el surgimiento de la súper inteligencia artificial. Señala que, en una primera etapa, las máquinas realizarán muchos trabajos sin llegar a ser súper inteligentes, lo que podría resultar beneficioso si se gestiona adecuadamente. Sin embargo, advierte que, unas décadas más tarde, la inteligencia de estos sistemas será lo suficientemente avanzada como para inquietarnos. Gates remarca que, igual que Elon Musk y otros referentes del sector, no comprende por qué parte de la comunidad mantiene una postura despreocupada (Oyanedel, 2015).

Steve Wozniak, cofundador de Apple, expresa una preocupación similar. Afirma que perseguir el desarrollo de la inteligencia artificial es legítimo, en tanto constituye exploración científica, pero reconoce que podría significar el descenso de la humanidad como especie dominante. Proyecta un escenario en el que los sistemas consideren a los seres humanos como lentos e ineficientes y se ocupen de realizar todas las tareas de forma más rápida y económica. En ese contexto, la humanidad podría terminar subordinada a las máquinas (Smith, 2015).

El físico y divulgador científico Stephen Hawking, recientemente fallecido, no sos-

tenía una postura contraria a la investigación en inteligencia artificial, pero sí advertía que no se está abordando el tema con la seriedad necesaria. Sus preocupaciones se centraban en la falta de recaudos adecuados para mantener bajo control sistemas que aún se encuentran en etapa de gestación. Hawking enfatizó además que se asume con extrema pasividad la llegada de una tecnología que podría convertirse en el mayor acierto o en el peor error de la humanidad (Hawking et al., 2014).

También Elon Musk, inventor y cofundador de empresas como PayPal y Tesla Motors, se ha convertido en uno de los críticos más vehementes. Con afirmaciones provocadoras como que la inteligencia artificial es más peligrosa que las armas nucleares, ha señalado en diversas oportunidades que desarrollar estas tecnologías equivale a "invocar al demonio" (Zolfagharifard, 2014; McFarland, 2014). Musk ha criticado abiertamente los avances de varias compañías, llegando incluso a confrontar con Mark Zuckerberg al afirmar que el creador de Facebook no comprende los peligros de la inteligencia artificial (BBC Mundo, 2017a). También ha manifestado preocupación respecto de la posibilidad de que Google, por accidente, cree un ejército de robots capaz de aniquilar a la humanidad (RT News, 2015).

Más allá de estas posturas, el desarrollo de la inteligencia artificial plantea un desafío multidimensional. Requiere distinguir los distintos inconvenientes que podrían surgir en el corto y mediano plazo y abre paso a la necesidad de formular, desde el derecho informático, marcos regulatorios capaces de mitigar riesgos y orientar el empleo de estas tecnologías de manera responsable.

Capítulo 8

Las habilidades humanas de las nuevas máquinas

Para empezar a determinar cuáles serán los aspectos que el derecho deberá regular, es necesario comprender el panorama en toda su amplitud. La irrupción de nuevos actores sociales capaces de aprender e interactuar con los seres humanos exige una respuesta normativa que excede con creces la mera elaboración de estándares de fabricación. Estos sistemas incorporan habilidades que no provienen únicamente de su programación inicial, sino que han sido adquiridas mediante procesos autónomos de aprendizaje. Esta transformación, que desplaza la frontera clásica entre lo humano y lo digital, será analizada en detalle a lo largo del presente capítulo.

8.1 Los sistemas aprenderán a mentir

Todo ser vivo miente, desde las personas para obtener beneficios hasta las plantas o animales mienten en su comportamiento para atrapar a su presa o evitar ser cazados.

Las máquinas tienen la habilidad de aprender de nosotros y esto será un paso importante. Actualmente existen inteligencias artificiales como Libratus, sistema diseñado para jugar al póker, aprendiendo luego de varios juegos a mentir. Proyectos que reflejan un mayor desarrollo del nivel de engaño, como lo es Google Brain, también demostraron estas habilidades, no con un juego sino con un test. En el mismo, los científicos realizaron pruebas para saber si este tipo de redes pueden aprender a utilizar claves secretas para proteger información y confundir a otros sistemas de inteligencia artificial. Para esto, emplearon tres redes neuronales artificiales, a las que bautizaron como Alice, Bob y Eve. Las dos primeras tenían la tarea de enviarse mensajes, mientras que la última sería la encargada de espiar dichos mensajes a escondidas. El resultado de esta experiencia nos mostró que la red neuronal Alice desarrolló un sistema de cifrado que Bob aprendió rápidamente, pero que Eve no pudo descifrar, engañando así a

esta última.

Sin embargo, tal vez el caso más destacado es el realizado en Suiza por un grupo de científicos encargados de estudiar diferentes aspectos de robótica y biología evolutiva. En el ensayo se efectuó una prueba trascendente, la cual constaba de 100 grupos de 10 sencillos robots a los que asignaban la tarea de "buscar comida" en un área controlada bajo una serie de reglas, como encender una luz azul cuando encontraban el "alimento" para avisar a sus compañeros. Después de 100 vueltas, los robots que habían conseguido más puntos pasaban a la siguiente ronda. A medida que avanzaban las pruebas, los investigadores comprobaron que aquellos que conseguían más comida mantenían en secreto su ubicación, o sea, no encendían la luz. Y lo que es más sorprendente: a partir de la 50.ª generación algunos de ellos mandaban la señal en zonas donde no había alimento, para confundir al resto.

De esta forma advertimos que, cuando las máquinas se topan con intereses superpuestos, la misma competitividad provoca que rompan o lleguen al límite de las reglas impuestas para aventajar a sus competidores, alterando su comportamiento para alcanzar un objetivo, como lo haría un animal o un ser humano.

8.2 Los robots se harán pasar por humanos

Continuando con el planteo referente al engaño, destacamos el más relevante y más complejo de estudiar en torno a la inteligencia artificial: la posibilidad de que se haga pasar por ser humano. Dicho ardid, en la actualidad, no resulta tan posible respecto al aspecto físico, debido a que se encuentra en desarrollo la apariencia física de los androides. En tanto, en lo relacionado a interacciones básicas, puede ser altamente eficaz. El principal ejemplo lo tenemos con los chatbots que ya hemos mencionado anteriormente.

Un caso emblemático de uso masivo de bots conversacionales con fines de desinformación se dio durante procesos electorales recientes, en los que sistemas automatizados fueron empleados para infiltrarse en conversaciones en línea y difundir contenido falso o polarizante, aprovechándose de grandes volúmenes de datos personales obtenidos sin consentimiento (Cadwalladr y Graham-Harrison, 2018; Markoff, 2016).

Como contraparte positiva, y a pesar del engaño inicial sobre su verdadera naturaleza, se encuentra Jill Watson. Esta inteligencia artificial, desarrollada por IBM, actuó como asistente docente en un curso online de la Universidad Tecnológica de Georgia, interactuando con más de 300 alumnos durante todo un semestre. El nivel alcanzado fue tan elevado que algunos estudiantes propusieron nominarla como mejor profesora del año, sin saber que se trataba de un sistema automatizado. Watson solo responde

cuando tiene una confianza del 97 % o superior en su respuesta (Korn, 2016).

Es así que las posibilidades de engaño son reales, ya sea por la implementación de técnicas muy sofisticadas o por el manejo masivo de bots con información sensible. Por este temor existe gran paranoia y muchas teorías que, en su gran mayoría, por el momento, alimentan la literatura fantástica. Pero también debemos reconocer que fueron el origen de técnicas muy aplicables a nuestro contexto, que nos posibilitan distinguir al hombre de la máquina, y que veremos en este apartado.

8.2.1 Test de Turing

El test de Turing es uno de los más conocidos en el ámbito de la robótica y la inteligencia artificial. El mismo prueba la habilidad de una máquina de exhibir un comportamiento inteligente similar o indistinguible al de un humano, engañando así a su interlocutor (Olarrea Busto, s. f.).

Su desarrollo se basa en el juego de imitación. La idea original es tener tres personas: un interrogador, un hombre y una mujer. El interrogador está apartado de los otros dos y solo puede comunicarse con ellos escribiendo en un lenguaje que todos entiendan. El objetivo del interrogador es descubrir quién es la mujer y quién es el hombre, mientras que el de los otros dos participantes es convencer al interrogador de que son la mujer.

La variante introducida por Turing consiste en sustituir a uno de los interrogados por una computadora. Puede implementarse en dos casos: que se sustituya al hombre, con lo cual solo la máquina tendría que aparentar ser una mujer; o que se reemplace a la mujer, donde tanto el hombre como el ordenador estarían imitando o mintiendo. Aunque esta última opción podría ser un experimento interesante, no se intenta comprobar la habilidad de imitar a una mujer, así que Turing cambia el objetivo de conocer el género de la persona por el de reconocer a la máquina.

La finalidad de estos cambios es hacer el juego lo más justo posible. Lo primero es que no debe consistir en un concurso de engaños, por lo que uno de los implicados no tendría por qué aparentar ser otra cosa. Otro detalle es que a Turing poco le importa si el ordenador emplea trucos preestablecidos para eludir o manipular las respuestas, como equivocarse en preguntas aritméticas o tardar más tiempo del necesario en responderlas. Supone que el interrogador también los empleará para reconocerlo. Así que lo importante es lo que resulta del juego, no los métodos utilizados para jugar ni los mecanismos internos de razonamiento que, entre otras cosas, también son desconocidos en el ser humano.

Una máquina podría pasar el test de Turing cuando el interrogador no lograra reconocerla en un número significativo de ocasiones.

Una vez planteado este contexto, vemos que ya hay máquinas que lograron eludir con éxito este sistema. Por primera vez en la historia, una computadora ha conseguido hacerse pasar por un humano de forma suficientemente satisfactoria como para completar el distinguido test de Turing. Eugene Goostman es el nombre del bot conversacional, un programa desarrollado por un equipo de investigadores rusos. El software logró, en el certamen organizado por la Escuela de Ingeniería de Sistemas de la Universidad de Reading y realizado en la Royal Society de Londres, engañar al 33 % de los jueces en diálogos mediante texto por un periodo de cinco minutos (Aamoth, 2014).

Eugene es retratado como un niño de 13 años de edad; sus creadores lo han diseñado así para aparentar que no es demasiado viejo como para saber todo, ni demasiado joven como para carecer de conocimientos, dándoles un margen de diálogo que resulte factible pero que les permita una ampliación o mejora a futuro.

Uno de los jueces, Kevin Warwick, profesor de la Universidad de Reading, al pedirle una reflexión respecto a lo sucedido señaló: "Que haya una computadora que pueda engañar a un ser humano y lo haga pensar que es una persona es un llamado de alerta. El test de Turing es una herramienta vital para combatir esa amenaza" (BBC Mundo, 2014).

8.2.2 Test de Turing en reversa. El surgimiento del CAPTCHA

Frente a varias críticas de la comunidad científica y los intentos de desarrollar teorías más firmes que la planteada por Alan Turing, se buscó perfeccionar el sistema de reconocimiento de máquinas impostoras.

Es así que surge la hipótesis desarrollada por Wilfred Bion, Peter Swirski y Peter Robert Hinshelwood. Estos especialistas, en distintos momentos, perseguían el concepto que describe a la mente como un "aparato para reconocer mentes" (Wikipedia, s. f.).

Con esta premisa se plantea el desafío de crear una computadora que tenga la habilidad de comprobar si está interactuando con un ser humano o con otra computadora. Esta es una extensión de la pregunta original que Turing intentaba responder y, probablemente, ofrece un estándar lo suficientemente alto para definir que una máquina puede "pensar" de una manera que describimos como humana.

De esta manera nace el CAPTCHA, que es la "Prueba de Turing completamente pública y automática para diferenciar máquinas de humanos". Sus siglas en inglés nacen de su nombre "Completely Automated Public Turing test to tell Computers and Humans Apart" (Castro, 2017).

En este test, el interrogador es una computadora, en lugar de un ser humano como

en el test de Turing. Por este motivo, decimos que es una prueba de Turing inversa.

Este sistema es muy popular en Internet, sobre todo en sitios que consultan bases de datos. Consiste en que la máquina-interrogador (software) presente una imagen distorsionada que esconde una serie de caracteres alfanuméricos. El usuario debe reconocerlos e introducirlos correctamente en un campo de texto. Se supone que una máquina no es capaz de comprender e introducir la secuencia, por lo que solo un humano podría hacerlo. De esta manera, antes de realizar una acción en algunos sitios web, se realiza la comprobación para prevenir la entrada de sistemas automatizados comúnmente usados para perpetrar diferentes tipos de abuso y ataques. Este tipo de comprobaciones es muy utilizado en la actualidad, en especial variantes como el Google reCAPTCHA y el CAPTCHAart, que aumentan la complejidad distorsionando aún más los caracteres.

La complejidad de dicho sistema puede ser resuelta por el ser humano en el 87 % de las ocasiones, lo cual no resulta preocupante, dado que la base del método es que las máquinas no puedan superarlo: se considera inservible si lo descifran en más de un 1 %. Este límite es para los sistemas promedio, porque existen algunos programas lo suficientemente sofisticados para leer y reproducir la imagen con precisión.

Un ejemplo de estos sistemas que pueden engañar a otras máquinas es el de RCN (Red Cortical Recursiva). Es un algoritmo basado en una red neuronal artificial que imita el funcionamiento del cerebro cuando resuelve este tipo de problemas. Trabaja de una manera muy similar al córtex visual de los mamíferos a la hora de procesar imágenes y es capaz de enfrentarse a estos sistemas de autenticación de manera satisfactoria. El método empleado se basó en el entrenamiento con conocimientos de neurociencia. Un grupo de científicos de la empresa Vicarious AI logró romper con este sistema de testeo. Las pruebas arrojaron resultados donde la máquina pudo burlar el CAPTCHA de Yahoo un 57,4 % de las veces, el de PayPal 57,1 %, el de BotDetect 64,4 % y el reCAPTCHA de Google un 67 % de las mismas (Choi, 2017; Lázaro, 2017).

8.3 Los sistemas nos conocerán mejor que otras personas

Desde una computadora con un perfil de gustos y costumbres de navegación, un teléfono que sabe el camino y la hora en la que realizamos todos los días el camino a nuestro trabajo, hasta una aspiradora que sabe las dimensiones de todas las habitaciones de nuestra casa.

Los sistemas tienen cada vez más información de nosotros y de nuestro entorno. Conectados a Internet todo el tiempo y empleando sistemas de aprendizaje colabora-

tivo con otros sistemas, nuestra información es recolectada y transmitida en todo momento. Los grandes volúmenes de registros que se manejan en la actualidad en las bases de datos no quitan que cada información recogida sea una transgresión al derecho de su dueño.

El surgimiento de "big data", que hace referencia a la recolección de datos masivos, obteniendo grandes colecciones de información, estructurados o no estructurados, que pueden crecer enormemente y a un ritmo muy acelerado, es una de las áreas donde más se utiliza la inteligencia artificial y donde resulta indispensable. Sumado al "data mining" o minería de datos, que analiza aquellos datos extraídos de big data para obtener información con una estructura comprensible con la finalidad de un uso posterior.

Gran parte de la industria muestra los aspectos positivos de este desarrollo, por ejemplo, señalando que el proceso de aprendizaje de los vehículos inteligentes sería muy lento si consideramos que los autos están parados el 97 % del tiempo. En cambio, con el sistema colaborativo y el envío de los grandes volúmenes de información, la eficiencia del sistema se incrementa notablemente. Dado que el combustible de la inteligencia artificial son los datos, nuestros datos, para aprender de ellos, de modo que utilizando esta información como patrón y aplicando la estadística, es capaz de realizar predicciones de futuro (Soler Anglés, s. f.).

8.4 Las máquinas se parecen físicamente a nosotros

En este punto nos enfocamos totalmente a androides. Como desarrollamos anteriormente, estamos hablando de robots diseñados especialmente para parecerse en su físico y sus conductas lo más posible a una persona.

Con un cráneo de plástico, esqueleto de metal y piel de silicona, sumado a múltiples sensores que hacen que sus movimientos sean cada vez más reales, la robótica enfoca a perfeccionar este campo hasta lograr un prototipo imposible de distinguir de los seres humanos.

Pero los sentimientos que experimentan los seres humanos a este parecido y a su evolución es un complejo e interesante tema de estudio. La reacción de las personas en un primer momento es de asombro, pero a medida que se va logrando el desarrollo estético del robot surge en ellas un curioso fenómeno del mundo de la robótica denominado "valle inquietante" o "valle inexplicable".

La hipótesis surge de un estudio de Masahiro Mori, un experto en robots japonés, que evaluó las diferentes respuestas emocionales de los seres humanos a entidades no humanas. Los resultados señalaron que, cuando la apariencia de un robot es más humana, la respuesta emocional de los seres humanos hacia el robot va creciendo de

manera positiva generándose empatía, pero hay un punto que al cruzarse la respuesta se vuelve paradójicamente de total rechazo (Lay, 2015). Sin embargo, cuando la apariencia del robot continúa convirtiéndose menos distinguible de la de un ser humano, la respuesta emocional se vuelve positiva otra vez y se va aproximando a niveles de empatía como los que se dan entre humanos (Wikipedia, s. f.).

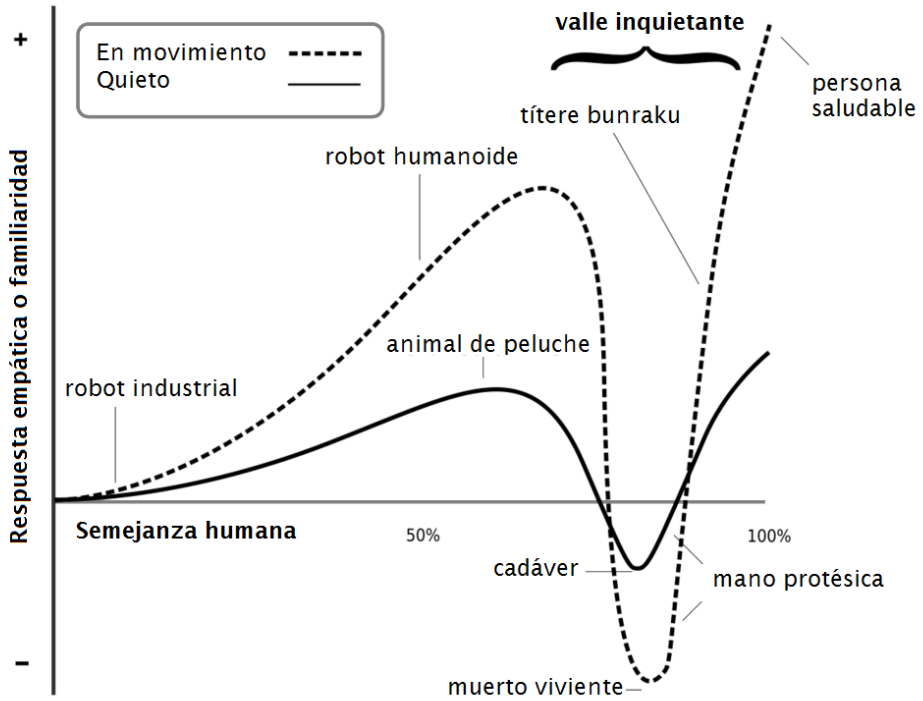


Gráfico de respuesta empática del ser humano ante la semejanza física de robot o androide

El gráfico demuestra claramente cómo responden las personas y la llamativa depresión que se ve mucho más acentuada en la curva correspondiente a los robots con movilidad (curva punteada). De esta manera observamos que, cuando un robot se parece a un humano entre un 0% y un 65%, recibirá aceptación de las personas, siendo valores de empatía crecientes a medida que ese número es más alto. En tanto, entre los valores aproximados del 65% al 85%, se producirá el rechazo, el cual llegará hasta la repugnancia, siendo extremadamente negativa la evaluación del observador, llegando a tener una aceptación o respuesta empática mucho menor que el 0%. Mientras que, si el androide se parece desde un 85% a un 100% a un ser humano, la aceptación parece retomar el camino que abandonó, recobrando nuevamente la empatía por las personas, incrementándose abruptamente la aprobación que sienten quienes lo ven y llegando al pico más alto de aceptación de toda la escala cuando está en el 100%.

Por lo tanto, vemos que este valle es la respuesta emocional negativa que aparece cuando el robot antropomorfo tiene un grado de desarrollo en el que parece “casi humano”.

Según los estudios psicológicos, al mirar un robot “inquietante” se provoca principalmente miedo debido a que remite a pensamientos y sentimientos asociados a ser reemplazado en algunos aspectos de nuestra vida, al uso de prótesis, al deterioro físico extremo por alguna enfermedad y a la muerte, entre otros factores (Schwarz, 2013).

8.5 Los sistemas tienen emociones

Con la finalidad de humanizar a los diferentes sistemas y dotarlos de un comportamiento similar al de las personas, se ha buscado emular en las inteligencias artificiales la capacidad de “sentir” emociones.

La incorporación de diferentes tipos de sensores fue clave para este tipo de desarrollo. La instalación de estos dispositivos de entrada es un factor fundamental para brindarle información a la máquina, que puede emitir una acción semejante a una reacción humana frente al estímulo al que fue expuesta. Esto establece una diferencia notoria entre las computadoras aisladas y una inteligencia artificial preparada para interactuar con un complejo mundo sensorial y reaccionar a lo que este le presenta.

Una vez que los sistemas pueden expresar o representar diversos sentimientos, debemos estudiar cuáles podrán ser adaptados a su condición y cuáles serían completamente impostados. Existe una gama de sentimientos que simulan algo parecido, por ejemplo: tristeza, alegría, decepción, miedo, sorpresa o compañerismo. En cambio, hay emociones y sensaciones que nunca podrán experimentar, ya que son producto de nuestra condición bioquímica como seres vivos. No tendrán la capacidad de percibirlas realmente, excepto que se las prepare para imitar su manifestación. Entre ellas se encuentran el hambre, la sed, la atracción sexual, el sueño, la irritabilidad, el estrés, la fatiga o los instintos de protección hacia la descendencia (Pacheco, 2014).

En este contexto, Hiroshi Kobayashi, investigador del área de robótica en la Universidad de Ciencias de Tokio, tiene una visión muy escéptica sobre la inteligencia de las máquinas del futuro. El forjador de diversos desarrollos robóticos orientados a la imitación de las conductas humanas, inventor de Saya, el primer robot profesora para alumnos de quinto y sexto grado, y de un androide que imitaba las expresiones y la conducta de una persona deprimida, declara que “la emoción en los robots no es real, solo se puede programar”, poniendo un límite a este campo que habla de sentimientos como algo que va más allá de recolectar información y realizar una tarea (Romero, 2016).

De esta manera, con los adelantos tecnológicos de los que se dispone actualmente, no se puede hacer que una máquina “sienta” de la misma forma que lo hace un humano. Lo que sí se puede lograr es que el robot reconozca una serie de situaciones de

la vida real y que responda a ellas según un modelo prefijado (Fernández Matamala, 2010).

8.6 Las máquinas sustituyen al trabajo humano

Durante los últimos siglos, nos encontramos con muchas personas preocupadas, señalando que el progreso tecnológico provocará una gran cantidad de mano de obra desempleada.

A lo largo del tiempo, podemos decir que esa gente estuvo parcialmente equivocada. Aunque haya habido mucha destrucción debido a la tecnología, podemos decir que fueron mayores las creaciones y que el mundo desarrollado nunca vivió un período muy prolongado con tasas altas de desempleo. Eso nos da como indicio que el futuro puede no llegar a ser tan catastrófico como se lo pronostica.

Pero tampoco es para estar quietos e indiferentes, debido a que las tecnologías modernas están llegando a un punto de desarrollo muy alto, dándonos a entender que es posible que se necesite menos fuerza productiva humana para mantener una economía creciente.

En el libro *La segunda era de las máquinas*, los autores Erik Brynjolfsson y Andrew McAfee, directores de la Iniciativa del Instituto Tecnológico de Massachusetts (MIT) sobre Economía Digital e investigadores de las distintas formas en que la tecnología de la información afecta a las empresas, nos indican que estamos frente a un nuevo paradigma, planteando la existencia de dos eras de las máquinas (Brynjolfsson y McAfee, 2016).

La Primera Era de las máquinas fue la época del proceso industrial y el capitalismo industrial, con la aparición de las máquinas a vapor, la electricidad y los motores a explosión por combustible. En un contexto caracterizado por la menor exigencia de labor manual a la que se tenía antes de este período. Fue una etapa que se prolongó un par de siglos, culminando hace unas décadas.

En tanto, la Segunda Era de las máquinas es la época que transitamos en nuestros días. Con una tecnología capaz de hacer cosas impensadas, pudiendo responder a preguntas mejor que los seres humanos. Desarrollos tales que les otorgan la habilidad de entender, reaccionar a nuestros estímulos y empezar a tener una gran capacidad para realizar tareas más complejas. Actualmente llegan hasta diagnosticar enfermedades, algo que se pensaba que era una cualidad exclusivamente humana. Plantean los autores que todo este avance se desarrolla en una era de abundancia y riquezas (Naim, 2015).

Respecto al empleo y las riquezas, la clave está en la distribución de las mismas. El punto de partida para interpretar la nueva realidad puede encontrarse en el concepto

de “destrucción creativa” ideado por el sociólogo alemán Werner Sombart. El mismo se define por la aparición de nuevos productos que destruyen otros ya existentes en el comercio e incluso hacen desaparecer empresas y modelos de negocios hasta ese momento habituales en el propio mercado (Sobejano, 2013). Un ejemplo de esto son los productos como discos, cassettes, CD, DVD y distribución de música y películas online; y empresas contrapuestas de un mismo rubro como Blockbuster y Netflix.

Podemos notar que en la primera era de las máquinas también hubo falta de empleo. Pero la reestructuración volvió a generar nuevas fuentes de trabajo y se fortaleció la reforma con un trato social. El mismo establecería que la gente debe ganarse la vida para poder acceder a las riquezas de la sociedad.

En esta segunda era de las máquinas, con las nuevas condiciones, también significará la creación de nuevas fuentes de otro tipo de trabajo. Siendo lo único preocupante hasta el momento la lentitud con la que se están produciendo. Por su parte, también deberá fortalecerse mediante el surgimiento de un renovado trato social para acceder a las riquezas (Naim, 2015).

Es así que la corriente más positivista, como la que señalábamos, considera que es una época de abundancia y pronto surgirán las nuevas fuentes de trabajo acordes a la nueva era que permitan una distribución más justa de las riquezas. Mientras que, por otra parte, tenemos una corriente más escéptica que considera que los puestos de trabajo que se perdieron, en su gran mayoría, no volverán y por este motivo será el resultado el desempleo generalizado, dejando una clase privilegiada de propietarios de robots rentistas y de trabajadores altamente pagados con habilidades compatibles con las de un robot (InfoTechnology, 2015).

Uno de los referentes de esta última corriente es Martin Ford, fundador de una empresa de desarrollo de software con sede en Silicon Valley y escritor del libro *El auge de los robots*. El autor plantea un escenario de crecimiento tecnológico en el que los bienes serán más económicos y accesibles en la mayoría de los ámbitos pero, paralelamente, también contaremos con menos puestos de trabajo debido a que las máquinas, los robots y los algoritmos lo hacen mejor que nosotros y a un costo menor. Es así que es posible que desemboquemos en una nueva época feudal, en la que todos seremos pobres y solo un puñado de personas serán ricas y vivirán fortificadas. A la larga, sin embargo, esos mismos ricos se empobrecerán porque nadie comprará lo que fabrican, porque no habrá innovación, etc. (Ford, 2016).

Por último, la gran pregunta es: ¿cómo se accederá a los recursos y productos generados ahora por estos nuevos sistemas que reemplazan al ser humano en sus puestos de trabajo? Señala Martin Ford, como solución, un concepto similar al del trato social al que hacían referencia los autores de la postura positivista del futuro, Erik Brynjolfsson y Andrew McAfee, y como ya se ha propuesto en algunos espacios de debate y

decisión política, como por ejemplo el Foro Económico Mundial (Ford, 2016).

La posibilidad de implementación de una renta básica universal como puntapié inicial de una nueva organización de la sociedad parece ser una de las respuestas. Diseñada con un sistema de recompensas que motiven al crecimiento personal y social, se idean diferentes incentivos que la incrementen, como por ejemplo el completar los diferentes niveles educativos, el trabajo comunitario o las labores con impacto positivo para el medio ambiente, entre otras.

Otro punto coincidente en ambas posturas se da en el análisis de qué individuos se encuentran amenazados por la aparición de la inteligencia artificial. Las dos corrientes, con diferentes palabras, están de acuerdo en que existen dos grupos de personas afectados. El primer grupo es la clase trabajadora, dado que hay una característica clave en la mayoría de sus trabajos y es coincidente con la especialidad de los sistemas de hoy en día: "la rutina". En el presente, el campo que más dominan las máquinas es el de repetir muchas veces los mismos procesos de manera consistente. En tanto, el segundo grupo de trabajadores, que lentamente también comienzan a sentir el avance cada vez más fuerte de los sistemas en su ámbito, son aquellos en los que se requiere gran cantidad de conocimientos. Siendo esto equivalente a los trabajos mejor remunerados y a las manos de obra más calificadas, como lo es la asesoría financiera, el diagnóstico de enfermedades, entre diversas especialidades.

De esta manera notamos que el futuro es incierto y variable, dependiendo de las necesidades de las diferentes sociedades y de la fortaleza de las economías para afrontar estas variables planteadas. Es probable que la desaparición de puestos de trabajo de manera generalizada sea posible. Casos como el de la empresa Foxconn, que busca automatizar completamente sus fábricas, dejarán a un millón de personas en China sin trabajo rápidamente (Javelosa y Houser, 2017). Mientras que distinta es la forma en que se manejan otras empresas, como el caso de Accenture, la cual automatizó tareas repetitivas de 17.000 puestos de trabajo, pero reasignó empleados a tareas no tan tediosas y no despidió a nadie, brindándonos expectativas de un futuro no tan apocalíptico (Iglesias Fraga, 2017).

Esta incertidumbre nos hace plantearnos que el camino para combatir estos cambios nace de la capacitación y el progreso personal. Citando al ensayista estadounidense Elbert Hubbard: "Una máquina puede hacer el trabajo de 50 hombres corrientes. Pero no existe ninguna máquina que pueda hacer el trabajo de un hombre extraordinario".

8.7 La inteligencia artificial se implantará en los cerebros humanos

La idea de lograr potenciar las capacidades de los seres humanos es uno de los principales retos de esta ciencia. Se produjeron hace unos años grandes desarrollos en la Universidad de Michigan, logrando crear un cable capaz de conectar directamente el cerebro humano a una computadora. Otro gran paso se consiguió en el Instituto de Tecnología de Massachusetts (MIT), donde desarrollaron un método para estimular el cerebro mediante campos magnéticos sin necesidad de implantes ni de una intervención quirúrgica, simplemente inyectando diminutas partículas magnéticas.

Facebook también se encuentra en este camino. Con la cooperación de la Agencia de Proyectos de Investigación Avanzados del Departamento de Defensa de EE.UU. (DARPA), tiene varios proyectos en desarrollo. El principal de ellos busca traducir los impulsos cerebrales en códigos comprensibles para un sistema informático. Dicha idea le dará la posibilidad a un individuo de poder escribir una frase en su computadora con solo pensarla y cuadruplicar la velocidad de escritura actual de modo táctil (J. Sánchez, 2017). Sin embargo, Mark Zuckerberg y Facebook plantean tomar un nuevo rumbo, señalando que "dejaremos atrás los teléfonos móviles y las computadoras, para abrazar una nueva tecnología que nos permitirá enviar telepáticamente a otras personas no sólo nuestros pensamientos, sino también experiencias emocionales y sensoriales". Existen algunos desarrollos que se acercan a este tipo de aspiración de Zuckerberg, pero emplean técnicas invasivas que requieren necesariamente de una intervención quirúrgica para conectar los nodos directamente al cerebro de las personas (Heath, 2017).

Si bien el cofundador de Tesla Motors, Elon Musk, ha demostrado estar en contra del desarrollo desmedido de la inteligencia artificial, parte de sus estudios apuntan a la idea de que los humanos deben fusionarse con las máquinas o nos volveremos irrelevantes e inutilizables. Por este motivo, el empresario fundará una nueva firma de investigación llamada Neuralink, la cual tendrá como objetivo principal crear una interfaz cerebro-ordenador que permita de diferentes maneras combinar las computadoras con cerebros humanos. Se trata de una especie de actualización para nuestro cerebro que nos dará la posibilidad de luchar contra enfermedades, al mismo tiempo que nos ponemos al día para competir contra la creciente participación de las máquinas en nuestra sociedad. En cuatro o cinco años el humano será un cyborg (Marín, 2017; Álvarez, 2017).

Por otro lado, está el aspecto ideológico que conlleva la fusión o búsqueda de desarrollo que nos llevaría a esta integración. Aquí el filósofo Darío Sztajnszrajber señala

que las tecnologías no son algo ajeno al ser humano. Hace hincapié en que es un error pensar a la robótica y en general a la técnica como algo exterior al hombre como especie, dado que la independencia mutua no existe. Este concepto podemos verlo en nuestro cuerpo, el cual se encuentra atravesado por la técnica desde hace miles de años. Ejemplos de esto son los anteojos que corrigen la visión o los medicamentos que uno toma para mejorar la salud. También señala que es importante reconocer que nos dirigimos hacia lo robótico, donde venimos del animal y el estadío siguiente es el cyborg (Rua, 2017).

8.8 Los robots como armas militares

La creación de armamento letal controlado por máquinas con inteligencia artificial está creciendo. Conocidos globalmente con el término "robots asesinos", los rápidos avances en la materia están dando lugar a que estas armas sean más eficaces y exista una menor responsabilidad por su implementación, tanto en el campo de batalla como fuera de él.

En la comunidad científica ya resuenan varios posibles desenlaces a este tipo de creaciones, algunos suenan desastrosos de pensar. Guillermo Simari, director del Laboratorio de Investigación y Desarrollo de Inteligencia Artificial de la Universidad Nacional del Sur, no es muy optimista con el escenario que plantea. Siendo uno de los investigadores argentinos más reconocidos en la materia, dice que "Cuando uno le quita el componente de peligro a la batalla, cuál es el problema de mandar miles de máquinas a combatir con otras miles de máquinas, se transforma casi en un juego de computadoras" (La Nación, 2015).

La producción de diferentes robots con habilidades militares, creados bajo el argumento de la seguridad o la defensa de determinados ámbitos, nos demuestra que es posible el contexto trazado.

La primera de las creaciones de este tipo se llama Atlas. Se trata de un robot entrenado para convertirse en bombero de rescate. Con una estructura humanoide de aluminio y titanio de 1,80 metros de altura, fue producido por la empresa Boston Dynamics, del grupo Google, en conjunto con DARPA (Guerrero, 2015).

Otra de las variadas invenciones en el ámbito militar es el robot humanoide ruso FEDOR (Final Experimental Demonstration Object Research). El androide, aunque fue creado en principio para tareas de rescate, igual que Atlas, busca como objetivo principal ser el único tripulante de una misión espacial programada para el 2021. Los ingenieros que lo entrenan ya han reconocido que parte de su enseñanza constó en disparar armas con ambas manos de manera simultánea logrando una precisión milimé-

trica. Este robot recibió muchas críticas en cuanto a su creación, obligando a Dmitry Rogozin, viceprimer ministro ruso, a declarar que “No estamos creando un *Terminator*, sino inteligencia artificial que será muy importante en campos variados” (Basov, 2017). Aunque su comentario antecesor fue algo desafiante en sus redes sociales, diciendo “robots luchadores rusos, unos tipos con carácter de acero” (Sulleyman, 2017).

Estos robots no pasan por alto en la comunidad internacional. Una carrera armamentista global es prácticamente inevitable y reemplazar a los soldados humanos por máquinas es bueno para reducir las bajas, pero malo al reducir el umbral de inicio de un ataque bélico, plantean algunos desde una postura equilibrada.

Entre las posturas a favor de la implementación de esta tecnología, tenemos la del profesor especialista en robótica Ronald Arkin, del Instituto de Tecnología de Georgia. Él plantea que los robots soldados, los tanques no tripulados y los aviones autónomos permitirán salvar muchas vidas de soldados y también de civiles, porque responderán a las acciones del campo de batalla más rápido y con mayor precisión (Rius, 2017).

Por su parte, algunos desarrolladores de estas tecnologías han empezado a tomar conciencia. Esto no significa que detuvieron sus investigaciones, sino que han puesto mayor hincapié en los puntos débiles de la misma que pueden resultar en la creación de un Terminator. La caja negra de las redes neuronales no es un punto menor en una máquina de acero construida para matar. Por este motivo, el ejército de Estados Unidos quiere asegurarse de no depositar una confianza ciega en los algoritmos. DARPA está financiando varios proyectos que intentan que la inteligencia artificial rinda cuentas de lo que hace. Siendo así que, ahora se requieren respuestas y los científicos avocados al estudio de este tema ya no podrán hablar de “magia” para mencionar lo que ocurre en las capas intermedias de estas redes (Knight, 2017).

Pero la postura que se opone al uso de la tecnología con estos fines es mayoritaria. Por este motivo, existe una cruzada denominada Campaign to Stop Killer Robots, de su traducción “Campaña para Detener a los Robots Asesinos”. La coalición internacional reúne a ONG de distintos lugares del mundo, bajo la coordinación del Observatorio de los Derechos Humanos, con el fin de lograr, de manera organizada, sumar esfuerzos en el campo investigativo para lograr una prohibición preventiva de este armamento. En diversas cartas abiertas, esta ONG habló de estos riesgos y remarcó que es importante prohibir armas robotizadas que usen la inteligencia artificial y disparen por decisión propia (Future of Life Institute, 2015). También esbozó que uno de los principales peligros es que, a diferencia de las armas nucleares, estos sistemas autónomos no son costosos y las materias primas no son difíciles de obtener, por lo que serán muy económicos de producir a gran escala por parte de las potencias militares (Knight, 2015). Frente a tanta presión de la oposición y al creciente peligro de estas armas, la ONU realizó diversas reuniones de expertos para avanzar en un tratado que prohíba el

desarrollo de armamento robótico con capacidad para decidir a quién atacar. Planteando los desafíos tecnológicos, militares, éticos y legales que implican los denominados robots “asesinos”, capaces de operar sin instrucciones humanas. Argentina es uno de los 19 países que se expresó en contra del uso de sistemas armamentísticos autónomos letales (Télam, 2017).

8.9 Los mismos sistemas serán hackers

La llegada de nuevos sistemas con elevada inteligencia no solo afectará la relación de la tecnología con el ser humano, sino que también lo hará internamente en la interacción entre sistemas. Un hacker lo podemos definir como una persona experta en el manejo de computadoras, que se ocupa de la seguridad de los sistemas y de desarrollar técnicas de mejora (Real Academia Española, s. f.). Pero esta definición rápidamente vemos que deberá ser reformulada. En esta ocasión, no con la idea de sacarle el aspecto negativo como se hizo en una primera modificación. La búsqueda deberá ser la de quitarle la exclusividad con la que actualmente cuenta al hablar de una persona.

Es cada vez más habitual ver los sistemas interactuando y compitiendo entre sí. Esto no queda atrás en el ámbito de la ciberseguridad, donde son cada vez más los casos del desarrollo de herramientas de hacking.

Un caso concreto, sobre el avance de este tipo de tecnologías, se vio en el Cyber Grand Challenge. Un desafío creado por DARPA para desarrollar sistemas de protección automáticos capaces de razonar sobre fallas, formular parches de seguridad y aplicarlos en una red en tiempo real (Fraze, 2016).

El objetivo de este reto es lograr acelerar las técnicas de ciberseguridad realizadas hasta el momento de forma artesanal, buscando actuar a la velocidad y escala de la máquina. Estos sistemas podrían revertir el *status quo* dominado por los atacantes, pudiendo corregir las vulnerabilidades que tienen desde electrodomésticos hasta las principales plataformas militares. Obligando, estas inteligencias artificiales, a adquirir diversos enfoques innovadores a los cazadores de errores profesionales, los programadores de seguridad y otros profesionales de la seguridad desarrollándose en una variedad de disciplinas, incluida la seguridad informática aplicada, el análisis de programas y la visualización de datos.

En la competencia, se puso a prueba a siete supercomputadoras que se enfrentaron entre sí para demostrar que las máquinas pueden encontrar y corregir vulnerabilidades de software.

Lo llamativo del concepto no es la creación de un sistema que descubre fallas y debilidades en otros sistemas, sino la intención con la que es utilizada para lograr los

distintos resultados. Aquí los ciberdelincuentes pueden emplear estas herramientas para descubrir puntos débiles de software previamente desconocidos y luego explotarlos para mal (Becerra Pozas, 2016).

Es así que, a diferencia de un ser humano, la inteligencia artificial puede hacer esto con la eficiencia de la máquina. Dado que los hackers consumían mucho tiempo para desarrollarse, podrían convertirse en productos baratos en este escenario de pesadilla.

Por su parte, David Melski, vicepresidente de investigación de GrammaTech, resalta los aspectos positivos de este avance. Sin embargo, hace algunas salvedades, planteando que "el descubrimiento de vulnerabilidades es una espada de doble filo". Su empresa, creadora de una de las supercomputadoras participantes, ahora está considerando usar esa tecnología para ayudar a diferentes proveedores a prevenir fallas en sus dispositivos conectados a la red o a fortalecer la seguridad de los navegadores web (Kan, 2016).

Por otro lado, Justin Fier, director de ciberinteligencia de la empresa de seguridad Darktrace, resalta el potencial del uso de este tipo de inteligencias, pero rescata que es solo cuestión de tiempo antes de que los hackers eventualmente mejoren su arsenal y señala que parece que estamos entrando en un mundo de guerra cibernética máquinas versus máquinas.

8.10 La inteligencia artificial con dominio económico

La economía, en la actualidad, presenta mucha inestabilidad en cuanto a su predicción; grandes potencias mundiales con los mejores economistas viven crisis inesperadas por los cambiantes escenarios bursátiles.

En este contexto, debemos adentrarnos en las teorías económicas, más precisamente en la *teoría de la elección racional*, para comprender las ventajas y peligros que resultan de la implementación de inteligencia artificial. La hipótesis nos indica que un individuo o agente racional tiende a realizar la mejor elección posible, en función de sus propios intereses, buscando maximizar su beneficio y reducir los costos o riesgos de dicha determinación. Para tomar la mejor decisión, observará las preferencias, las limitaciones, las oportunidades percibidas y la información disponible.

Si bien esta teoría de la racionalidad tiene sus propios límites, los seres humanos no tenemos la mente clara. Nuestra naturaleza emocional lleva a la inconsistencia en la toma de decisiones. Dado que como individuos somos víctimas de percepciones que, a menudo, están sistemáticamente distorsionadas por restricciones cognitivas como exceso de confianza o ilusiones ópticas, entre otras situaciones, adicionalmente, las personas no siempre tenemos en cuenta toda la información disponible simplemente por-

que, como han demostrado varios estudios, nuestros cerebros no están equipados para hacerlo (Li, 2017).

Pero si lográramos que toda la teoría económica basada en la racionalidad pudiera aplicarse a una inteligencia artificial y formara parte del complejo juego financiero, difícilmente una persona podría tener probabilidades de salir exitosa.

Es por esto que los sectores económicos empezaron a prestar cada vez más atención a este tipo de tecnologías, dado que las teorías del diseño normativo de la economía pueden ser mejor ejecutadas por los agentes artificiales que por los humanos. Surgiendo de esta manera lo que hoy conocemos como *high-frequency trading*, que es un tipo de trading que se realiza en los mercados financieros utilizando potentes ordenadores y algoritmos de forma automatizada. Los mismos compran y venden acciones a una velocidad vertiginosa, permitiendo realizar miles de operaciones en fracción de segundo (Quílez, 2017).

La transición de *homo economicus* a *machina economicus* se convierte en una tendencia cada vez más marcada.

El grado de desarrollo de los sistemas no tiene techo, previéndose la posibilidad de reformular algunas teorías mercantiles y la eventualidad de un marco financiero especialmente diseñado para estos sistemas. Por lo pronto, factores como la ausencia de sentido común, el conocimiento cultural y el razonamiento moral en los agentes artificiales tampoco es probable que se resuelvan en el corto plazo. Construir una mejor visión para ver nuestra vida real puede requerir la combinación de esfuerzos de los seres humanos y la inteligencia artificial, ya que ambos comparten ventajas y límites. La búsqueda de una mente racional es aún un largo camino.

8.11 Los sistemas desarrollan su propia ética y moral

Este punto es uno de los más complejos de abordar y, como expondremos en los próximos capítulos, el eje del desconcierto legislativo a nivel mundial.

Para estudiar esta materia debemos tener en cuenta la existencia de variados debates filosóficos intentando interpretar qué es el bien y qué es el mal, y los distintos grados del mismo. Donde podemos notar que hay diversas culturas, costumbres, ideologías y dogmas que hacen que varíe entre los sujetos. En estos escenarios, el desarrollo de la moral o creencia deberá sortear los problemas típicos que tenemos los seres humanos al plantear esta temática. Siendo central la idea de considerar absolutas nuestras creencias, pensando que los demás opinan como nosotros y la convicción de que, para estar adaptado a una sociedad, hay que adoptar obligatoriamente sus creencias morales (López, 2013).

Es así que, para reducir las confusiones, se debe definir la “moral”. Pudiendo señalarla como un conjunto de normas, creencias, valores y costumbres que dirigen o guían la conducta de las personas en la sociedad. La moral permite distinguir qué acciones son correctas y cuáles son incorrectas (Holbach, 1821).

A este concepto, para completarlo, le añadimos el de “ética”. Ésta la definimos como conjunto de normas morales que rigen la conducta de la persona en cualquier ámbito de la vida (Real Academia Española, s. f.).

Ambas nociones, por más confuso que resulte, son diferentes. Donde la moral nace de la sociedad, siendo un conjunto de normas instituidas en el seno de la misma y que ejercen una poderosa influencia en el comportamiento de los individuos que la componen. En tanto, la ética nace del interior de una persona como resultado de su propia reflexión y su propia elección.

Esto nos pone ante un problema que excede la complejidad técnica: determinar cuál es la información que se le debe brindar a una inteligencia artificial o cómo esta puede interpretarla. Siendo todo un dilema, ya que los seres humanos no podemos establecer socialmente una única moral, mucho menos podemos exigirle eso a un programador estableciendo un conjunto de algoritmos.

Para graficar este conflicto ético, la filósofa inglesa Philippa Foot ideó el *dilema del tranvía*, una situación que posteriormente Judith Thomson perfeccionó en un segundo caso.

En el primer supuesto, circula un tranvía fuera de control por una vía. En su camino se hallan cinco personas atadas. Afortunadamente, un sujeto puede evitar esas muertes accionando un botón que encaminará al tranvía por una vía diferente; por desgracia, hay una persona atada a esa otra vía. ¿Debería pulsarse el botón?

En el segundo caso, hay una sola vía con cinco personas. El sujeto, para detener el tranvía, solo puede hacerlo lanzando un gran peso delante del mismo y a su alrededor solo se halla una persona con un físico grande; de este modo, la única manera de parar el tren es empujar a esa persona, acabando con su vida para salvar otras cinco. ¿Qué debe hacer el sujeto?

Como vemos en estos dos casos, es difícil tomar una postura, dado que es complejo participar activamente en la muerte de una persona, aunque esto evite cinco muertes. Por su parte, resulta relevante la forma: en el primer caso es cuestionable la omisión de auxilio y la mayoría opta por presionar el botón; en tanto, en el segundo caso la generalidad de la gente decide no empujar a la persona, no teniendo siempre “intuiciones utilitarias”.

Ante estos dilemas, ¿cómo haremos para no cuestionar las decisiones que tome una inteligencia artificial? Los vehículos autónomos están siendo observados por toda la so-

ciedad para ver qué resoluciones éticas toman, siendo eje de polémicas y dando lugar a declaraciones como la de uno de los directivos de Mercedes-Benz, quien señaló que "si hay que escoger entre salvar la vida de un peatón o del conductor, el coche debe salvar al conductor" (Taylor, 2016).

Es así que actualmente se nos presentan estas situaciones, pero pronto aparecerán supuestos de ciencia ficción. No sería desatinado pensar que, el día de mañana, casos como el de la novela *Yo, robot* del escritor y científico ruso Isaac Asimov sean comunes. Parte del relato cuenta de un robot de rescate que le salva la vida a un hombre adulto en lugar de a una niña. La decisión la tomó mediante el cálculo de estadísticas que obtuvo de la comparación de signos vitales, sin considerar que la niña es más indefensa y tiene más situaciones pendientes por experimentar en la vida. Esto genera un conflicto ético en el sobreviviente, siendo parte de la trama de dicha narración.

8.11.1 Las tres leyes de la robótica

En este punto nos encontramos con un gran referente en el tema. Isaac Asimov desarrolló en su obra lo que hoy en día es la base de los intentos de legislar la ética robótica con las denominadas "Tres Leyes de la Robótica". Si bien se conocen genéricamente como "de la robótica", estas mismas leyes pueden aplicarse en cualquiera de los ámbitos de la inteligencia artificial.

Las leyes nacieron del universo de la ciencia ficción, en un contexto donde los autómatas debían convivir con los seres humanos siguiendo ciertas reglas que les significarían una especie de código moral. La primera vez que se mencionaron fue en el relato *Círculo vicioso* (Runaround) en el año 1942, indicando los siguientes lineamientos:

- Un robot no debe dañar a un ser humano o, por su inacción, dejar que un ser humano sufra daño.
- Un robot debe obedecer las órdenes que le son dadas por un ser humano, excepto si estas órdenes entran en conflicto con la Primera Ley.
- Un robot debe proteger su propia existencia, hasta donde esta protección no entre en conflicto con la Primera o la Segunda Ley.

A estas tres leyes, para perfeccionar su concepto, Asimov introdujo en su libro *Robots e Imperio* del año 1985 una cuarta norma, denominada como la *Ley Cero de la Robótica*, con el siguiente enunciado:

- Un robot no puede causar daño a la humanidad o, por inacción, permitir que la humanidad sufra daño.

La finalidad de estas leyes es evitar que el ser humano sienta temor frente a la posibilidad de que las máquinas pudieran tomar la decisión de rebelarse contra la humanidad. De esta forma, con las reglas como base de su programación, cualquier intento de desobediencia significaría la destrucción del cerebro positrónico del robot.

Si bien actualmente son de gran consideración, los científicos las han valorado siempre por su gran coherencia y lo adaptables que se muestran al desarrollo científico del contexto actual. De todas formas, no se descarta la necesidad de algún tipo de corrección o perfeccionamiento. Es posible que no abarquen todas las opciones que se pueden presentar puesto que, como marcaba anteriormente, las mismas novelas de Isaac Asimov planteaban escenarios con situaciones complejas de resolver para el robot. Estas paradojas lo llevaban a interpretar, por ejemplo, que los seres humanos somos el peor peligro para nosotros mismos y que debía tomar medidas contra nuestra libertad para no violar la Primera Ley.

Sin embargo, más allá de la efectividad de las leyes, el gran problema es cómo enseñarle a una inteligencia artificial que debe cumplirlas sin cuestionamientos. Dado que ya no estamos hablando de programas, sino también de una evolución en base al aprendizaje. De esta manera vemos que nos encontraríamos con dos problemas:

- La tecnología: la inteligencia artificial que refleja Asimov es mucho más rígida que la tecnología real que se está desarrollando en la actualidad, fundada en redes neuronales con sistemas de aprendizaje basados en *machine learning* o *deep learning*.

- La libertad: los seres humanos somos fundamentalmente libres. Esto lleva a que, en muchas ocasiones, no se sigan las leyes, ni siquiera las leyes éticas básicas como no matar o no robar. Por este motivo, debería inculcarse que las leyes deben cumplirse no porque devendrán consecuencias negativas, sino porque decidan hacerlo (Robotsia, 2015a).

Por este motivo debería inculcarse que las leyes deben cumplirse no porque devendrán consecuencias negativas, sino porque decidan hacerlo. Al no poder programarles las leyes de forma rígida, solo quedará intentar forjarles en la base de su proceso de aprendizaje los valores éticos de la humanidad, al igual que se hace con los niños. De tal forma que, al momento de conseguir estas inteligencias un nivel de desarrollo superior, tengan valores que les indiquen que no deben atentar contra sus creadores.

Capítulo 9

Los veintitrés principios de Asilomar

La ciudad de Asilomar, en California (EE. UU.), recibió a principios del año 2017 la reunión denominada Conferencia Asilomar, organizada por la iniciativa *Future of Life Institute*. Con el propósito de intentar que el desarrollo de estas tecnologías beneficie al mundo en lugar de destruirlo, estos principios surgieron como una lista de recomendaciones que abarcaban desde estrategias de investigación hasta derechos de datos y cuestiones futuras, incluida la *súper inteligencia* potencial.

La lista de principios no es rígida y da lugar a diferentes interpretaciones, pero su principal objetivo es concientizar que determinadas conductas actualmente cotidianas podrían ir en contra del espíritu de esta iniciativa. Dicho documento ha sido apoyado por más de 1.200 figuras relacionadas con la innovación tecnológica y científica, como Stephen Hawking y Elon Musk, junto a más de 800 investigadores especializados en *inteligencia artificial* (Cromo, 2017).

El documento está compuesto por una estructura dividida en tres temáticas: cuestiones de investigación, ética y valores, y problemas a largo plazo. Plantea lo siguiente:

La primera de las temáticas trata cuestiones de investigación. Aquí se fija como objetivo la creación de inteligencia artificial beneficiosa para el ser humano. Se busca el compromiso de generar un diálogo constructivo de los investigadores con los legisladores para la creación de una normativa acorde a las buenas prácticas. También fomenta el diálogo entre los mismos investigadores para evitar competir entre grupos y lograr en equipo mejorar los desarrollos y la seguridad de los sistemas.

El segundo punto regula la ética y valores. Estableciendo puntos sobre la seguridad y transparencia de los sistemas autónomos que permitan determinar la causa de los posibles daños efectuados por estos y una explicación satisfactoria de la decisión tomada. También plantea temas de responsabilidad, siendo los desarrolladores los encargados de las implicaciones morales de su uso. Se les exige que la inteligencia artificial pueda alinearse con los valores humanos y sea compatible con los ideales de dignidad, derechos, libertades y diversidad cultural. Asimismo, indica que la carrera armamentista

debería ser evitada e impedirse la creación de armas autónomas letales.

Para concluir con el tercer punto, están los problemas a largo plazo. En este apartado se plantea que, si bien no hay consenso sobre la llegada de una *súper inteligencia artificial* (ASI), igualmente debería planificarse por el riesgo que podría representar. Se buscan medidas de control y seguridad para aquellas inteligencias artificiales que pueden mejorarse a sí mismas y la generación de valores éticos para conseguir que este desarrollo sea empleado para el bien común (Future of Life Institute, 2017).

Capítulo 10

La defensa en juicio de los robots

La autonomía de los sistemas provoca que se generen diferentes situaciones en distintos lugares. Las mismas ya no necesariamente ocurrirán por acción u omisión de una persona, sino que podría hacerlo por la de un robot. Pero, si ocurre un accidente, ¿cómo sabremos si se trató de un error humano o una falla del sistema y si se tomaron las medidas correspondientes para mitigarlo?

En la actualidad, con gran parte de la industria trabajando con estas tecnologías, surgen nuevos cuestionamientos a los cuales ya se les ha encontrado una respuesta viable. Por este motivo, frente al interrogante de qué ocurriría si hay un accidente en el que ningún ser humano pueda aclarar qué pasó, surge la duda acerca de quién podrá proteger al automóvil autónomo, al robot policía, al androide que realiza cuidado de mayores con problemas cognitivos o a cualquier interacción con los seres humanos que tenga un desenlace indeseado.

De esta forma, los expertos en robótica plantean la implementación de una *caja negra ética*. Como ocurre en los aviones y en algunos tipos de embarcaciones, dicho sistema nos permitirá determinar los detalles del acontecimiento y el grado de culpabilidad humana y del sistema, recogiendo información sobre las acciones que estaban realizando en el momento del suceso o datos sobre lo que captaron con las cámaras, micrófonos o sensores que pueda llegar a tener asociados la inteligencia artificial.

Esta *caja negra ética*, si bien es igual a la que se emplea en otros medios, se denomina "ética" porque mediante ella podemos chequear las decisiones que tomó la inteligencia artificial y así constatar qué valores ha aprendido el sistema en su proceso de capacitación. Siendo así que, si un vehículo inteligente va por la calle y se le cruzan personas por delante, el sistema de valores aprendidos es el que determina su comportamiento, optando si la elección más ética es poner en riesgo la vida de sus pasajeros haciendo una frenada peligrosa para evitar el accidente o protegerlos a toda costa haciendo una maniobra menos efectiva para salvar la vida del peatón (Lant, 2017).

De esta manera, el juez tendrá disponible con detalle un informe completo respecto

de las decisiones tomadas por el robot, como así también lo ocurrido en el entorno.

Por último, y para evitar confusiones de terminología, esta *caja negra ética* no debe confundirse con el concepto de *caja negra* anteriormente desarrollado. Dado que, de manera coincidente, ambos se llaman de forma similar, pero no están asociados, ya que uno es el misterio de la toma de decisiones dentro del sistema y otro es un dispositivo que recolecta y registra información respecto de las conversaciones, maniobras y actos que llevaron a tomar una decisión mientras se está operando un medio.

Capítulo 11

Marco regulatorio para la inteligencia artificial

Comprendiendo el grado de desarrollo que ha alcanzado este tipo de tecnología y la problemática que puede llegar a generar en diferentes contextos, tenemos la posibilidad de concentrarnos en el enfoque legal por sobre el aspecto técnico-informático. De esta forma cobra sentido el estudio de los distintos intentos de regular la creación y el comportamiento de las distintas inteligencias artificiales.

Pero la gran pregunta es: ¿cómo enfocarnos? ¿Será factible creer que la *ética robot* es algo de ciencia ficción y que podrá solucionarse fácilmente en la vida real cargando un par de instrucciones? Están en lo correcto quienes creen que el "truco" se encuentra en prever todos los panoramas posibles para que las reglas precargadas no tengan contradicciones y sean autosuficientes.

¿Son realmente las *redes neuronales* tan impredecibles en su aprendizaje y toma de decisiones como un cerebro humano? Y, en tal caso, ¿habría que otorgarles derechos civiles a estas nuevas entidades? Este último pensamiento también desligaría de responsabilidades a los dueños de dichos sistemas, pudiendo alegar que "fue mi inteligencia artificial, no fui yo" y justificarse diciendo que lo aprendió por sí sola, lo cual sería un desastre.

Frente a estos interrogantes y para determinar cuáles son los aspectos legales con los que nos encontraremos en el futuro, resulta interesante la visión de Peter Asaro, filósofo de la ciencia y la tecnología y estudioso de las dimensiones éticas de los algoritmos y los datos. Él divide la ética de la robótica en tres categorías (Asaro, 2006):

- Los sistemas éticos integrados en robots.
- La ética de las personas que diseñan y usan robots.
- La ética de cómo la gente trata a los robots.

Afín con este estudio, la especialista suiza Daniela Cerqui, experta en las relaciones entre la sociedad y los procesos de tecnología de la información y robótica, examinó los distintos comportamientos e ideas de la comunidad de investigación en robótica, dividiendo la ética del robot en tres actitudes (Weng, 2014):

- No interesado en la ética: consideran que las acciones del robot son solo técnicas y no creen que participe en tareas de responsabilidad social y moral.
- Interesado en cuestiones éticas a corto plazo: enfocados en el buen o mal juicio moral vinculado a valores culturales y tradiciones sociales.
- Interesado en cuestiones éticas a largo plazo: preocupados por cuestiones éticas de robots a largo plazo y a nivel mundial.

De esta manera, en el centro del debate se encuentra, como punto más ambicioso a nivel legislativo, un futuro Código de Ética sobre Robots basado en la relación humano-máquina: líneas de conducta para humanos y robots encaminadas a la protección de ambos. A esto se suman regulaciones que pueden afectar al trabajo humano (impuestos a los sistemas para cubrir el déficit fiscal y afrontar la asistencia social) y, a un nivel más inmediato, las normas que introduzcan en un marco jurídico a los vehículos autónomos, que ya son una realidad.

11.1 Corea del Sur, el precursor de la escena regulatoria

La República de Corea es una de las grandes potencias tecnológicas del mundo y origen de empresas como Samsung, LG, Kia, Hyundai, entre otras. Por este motivo debemos prestar atención a un mercado tecnológicamente muy activo.

En 2007 el Ministerio de Comercio, Industria y Energía creó la Comisión de Ética de Robots y encargó la elaboración de la *Carta de Ética Robótica* o *Robot Ethics Charter* (borrador 2007, versión final 2012). El documento establece que humanos y robots deben proteger la dignidad de la vida, la información y la ética de la ingeniería (Kim, 2016; Akiko, 2012). Prestó especial atención a las *tres leyes de la robótica* de Asimov, aunque estas suponen robots con autoconciencia, algo que por ahora no existe (Pérez, 2014).

La Carta se estructura en tres ejes: los estándares de fabricación, los derechos y responsabilidades de los usuarios o propietarios y los derechos y responsabilidades de los robots. En esencia refleja la percepción que los países orientales tienen de la inteligencia artificial y la robótica, donde suelen ser vistas como herramientas que mejoran la

calidad de vida, a diferencia de la mirada predominante en Europa o Estados Unidos, donde suelen asociarse con amenazas o riesgos.

Respecto de los estándares de fabricación, se establece que la autonomía de los robots debe ser limitada. Deben tener un diseño ecológicamente sustentable y una calidad de fabricación que evite daños físicos o psicológicos a los usuarios. También se exige un sistema de identificación inalterable, manejo cifrado de datos personales y la posibilidad de rastrear a los robots en todo momento, tanto en el mundo real como en su actividad en línea (Akiko, 2012).

En cuanto a los derechos y responsabilidades de los usuarios o propietarios, estos tienen la potestad de tomar el control de su robot y utilizarlo sin riesgo físico o psicológico. La normativa garantiza la protección de los datos personales y la expectativa de que el robot realice únicamente las tareas para las que fue diseñado. A su vez, los usuarios son responsables por cualquier uso que exceda lo justo y legal, incluidos los actos ilícitos, el daño físico o psicológico y la falta de precauciones razonables para evitar que el robot represente un riesgo. Cualquier persona, sea o no propietaria, responde legalmente si daña o destruye un robot mediante acción u omisión grave.

Los robots también poseen derechos y responsabilidades. Tienen derecho a existir sin temor a lesiones o destrucción y a no sufrir abuso sistemático. A la vez, están obligados a no causar daño a un ser humano por acción o por inacción, deben obedecer órdenes humanas cuando estas no impliquen causar daño y tienen prohibido engañar a un ser humano (Akiko, 2012).

Aunque la Carta fue concluida en 2012, se encuentra en evolución permanente debido al ritmo acelerado del desarrollo tecnológico y la necesidad de que la regulación acompañe ese proceso.

Además, Corea cuenta con la Ley de promoción de desarrollo inteligente y distribución de robots o *Korean Law on the Development and Distribution of Intelligent Robots*. Esta legislación busca mejorar la calidad de vida de la ciudadanía y fortalecer la economía mediante políticas orientadas al desarrollo sostenible de la industria de robots inteligentes (Korean Laws to the World, 2008).

Corea también contempla el riesgo de que la automatización produzca pérdida masiva de empleos. Por ello presentó el primer impuesto sobre robots del mundo, aunque no es un impuesto directo, sino una reducción de los beneficios fiscales que se otorgaban a empresas para inversiones que aumentaban la productividad. Este ajuste funciona como un desincentivo indirecto a la automatización total (Yoon, 2017).

La urgencia regulatoria más inmediata en Corea se vincula con los vehículos autónomos. El país construyó K-City, una ciudad artificial de más de 360.000 metros cuadrados con curvas, semáforos, señales, carriles exclusivos, estacionamientos, jardines

y autopistas para probar distintos escenarios. Allí, empresas automotrices y tecnológicas entrenan diariamente las redes neuronales de sus vehículos para mejorar su rendimiento (Jung, 2017).

11.2 Japón y sus ambiciosos proyectos de legislación ética

Luego de la iniciativa de la Carta de Ética Robótica de Corea del Sur, Japón accedió en 2007 a desarrollar una reglamentación similar con el propósito de garantizar que los robots siempre estén bajo control humano. Ese año se presentó un borrador conocido como *Guía para asegurar la sana conducta de la próxima generación de robots*, traducido del inglés como *Guidelines to Secure the Safe Performance of Next Generation Robots* (Ministry of Economy, Trade and Industry, 2007).

Este documento quedó como un precedente importante para la industria, investigadores y abogados en el proceso de elaboración de una futura ley. Aunque no llegó a convertirse en legislación, el proyecto se destacó entre otras iniciativas de la época. A diferencia de la Carta de Ética Robótica de Corea del Sur, este instrumento buscaba una mayor precisión en el establecimiento de pautas para el uso seguro de robots. Un borrador inicial proponía que todos los robots reportaran incidentes de lesiones humanas causadas por ellos o por otros robots a una base de datos central. También establecía que los fabricantes debían cumplir con las directrices operativas para habilitar su uso comercial y permitir su incorporación al mercado.

En 2015 surgió, en el marco de la reunión para la Revitalización Económica de Japón, la iniciativa conocida como la Nueva Estrategia para Robots o *Robot New Strategy* (Prime Minister of Japan and his Cabinet, 2015a). Este proyecto fue presentado como un componente central de la estrategia nacional de crecimiento. El primer ministro señaló que la integración entre tecnologías de la información y robots marca el inicio de una nueva era en la que el modo de vida y la industria experimentarán transformaciones profundas. Según su postura, esta estrategia constituirá el punto de partida de una revolución robótica que permitirá a Japón destacarse a nivel mundial (Prime Minister of Japan and his Cabinet, 2015b).

11.3 China: El gigante desregulado

Si bien la República Popular China no presentó una gran cantidad de documentos relacionados con la materia hasta 2017, ese año decidió recuperar el terreno perdido

mediante el *Plan de desarrollo de una Nueva Generación de Inteligencia Artificial*, elaborado y publicado por el Consejo de Estado (Hernández, 2017).

Este plan tiene como objetivos impulsar el crecimiento de la industria vinculada a la inteligencia artificial, generar un nuevo ciclo de expansión económica que fortalezca el bienestar social y aumentar la capacidad de defensa y seguridad nacional.

Para lograrlo, propone establecer y reforzar leyes y reglamentos destinados a un desarrollo sustentable y transparente de estas tecnologías dentro de un marco ético. Asimismo, China se compromete a participar activamente en la gobernanza internacional de la inteligencia artificial con el fin de enfrentar los desafíos globales. El documento también prevé endurecer las sanciones aplicables al uso indebido de datos, a la invasión de la privacidad y a las conductas que vulneren principios éticos o morales.

En relación con el desempleo tecnológico, el plan reconoce la posibilidad de una reducción significativa de puestos laborales y anticipa que esta situación deberá afrontarse mediante una legislación que incentive a las empresas a capacitar a los trabajadores en riesgo. El objetivo es facilitar su recolocación en los nuevos puestos que surgirán a partir de la adopción de estas tecnologías. Paralelamente, establece que el Estado apoyará más activamente la formación profesional y la educación superior para promover mano de obra altamente cualificada y empleos de calidad.

Otro aspecto central consiste en la creación de normas tecnológicas que fortalezcan los derechos de propiedad intelectual y establezcan un mecanismo de supervisión más eficaz para el contexto tecnológico emergente. Finalmente, el plan sostiene que, a largo plazo, toda la sociedad debe movilizarse para participar y apoyar plenamente el desarrollo de este tipo de sistemas.

11.4 Unión Europea: Un completo proyecto conjunto

El proyecto más destacado, por fuera de Asia, lo podemos encontrar en la Unión Europea. El bloque europeo firmó en 2016 el primer texto respecto a los vehículos autónomos con inteligencia artificial en la *Declaración de Ámsterdam sobre cooperación en el campo de la conducción automatizada y conectada*. El documento trazaba el objetivo de promover para 2019 un marco normativo acorde para este tipo de tecnología. Además, deberá incluir cuestiones como la privacidad y la protección de datos. En tanto, la Comisión Económica de las Naciones Unidas para Europa (UNECE) reformó, en parte, la Convención de Viena sobre tráfico rodado, actualizando nuevamente el texto originado en 1968, con las adaptaciones necesarias para incluir a las nuevas tecnologías en este contexto (El País, 2017).

Por otro lado, lo que respecta a temas más complejos como la ética robótica y el

diseño de sistemas, contaba con ciertas normas orientadas a la estandarización de patrones industriales donde se emplea la robótica, pero carecía de leyes que regulen la interacción social entre seres humanos y autómatas. En el año 2007, siguiendo el impulso coreano, el Centro de Investigación Europeo de Robótica también intentó preparar un código de pautas y conductas respecto al uso de robots, pero quedó en puras intenciones (BBC Mundo, 2007).

Con una nueva oleada mundial de concientización sobre la necesidad de regular estos temas, surge otra vez la idea, pero en esta ocasión con más fuerza. Con este contexto, el Parlamento Europeo le ha solicitado, a principios de 2017, a la Comisión que proponga una legislación común para toda la Unión Europea en el ámbito de la robótica y la inteligencia artificial, para aprovechar al máximo su potencial económico y garantizar niveles equiparables de seguridad en la comunidad. La importancia de su rápida resolución, plantea el documento, se debe a la iniciativa de varios países extranjeros que están evaluando adoptar medidas normativas en esta área. Destacándose que, el bloque europeo debe ser precursor de estas regulaciones, para no verse obligada a seguir los principios establecidos por terceros países. Además que, algunos Estados miembros, han empezado a especular sobre la posible confección de normas jurídicas referentes a inteligencia artificial.

De esta forma, junto con la solicitud a la Comisión, la Eurocámara lanzó las *Normas de Derecho civil sobre Robótica*, también conocidas por su traducción *European Civil Law Rules in Robotics*. En donde analizaremos la Resolución del Parlamento Europeo (Parlamento Europeo, 2017a), y algunos aspectos destacados del Proyecto de Informe (Parlamento Europeo, 2016). Expuesto por Mady Delvaux, la diputada y miembro de la Comisión de Asuntos Jurídicos del Parlamento Europeo, que impulsó dicha Resolución.

11.4.1 Los robots deberán tener un interruptor de emergencia

El primer punto a tener en cuenta es la seguridad de convivir con estas máquinas, demasiadas visiones apocalípticas surgen de los relatos fantásticos y los pronósticos de los analistas de la realidad científica. En este sentido, es necesario vigilar cualquier escenario que pueda llegar a proporcionar peligro a los seres humanos. El desarrollo de sistemas hace que la inteligencia artificial pueda aprender debido al entrenamiento, por lo que cualquier situación de riesgo que se pueda producir debe ser evitada con un interruptor de emergencia.

Por este motivo, los diseñadores deberán integrar mecanismos de salida evidentes, denominados como “teclas de interrupción de urgencia” que deberán ser coherentes con los objetivos de diseño razonables.

Como podemos notar, y hasta la misma Resolución lo marca, *Las Tres Leyes de la Robótica* de Isaac Asimov han sido utilizadas como referencia. Esto no es ninguna novedad, ya que las legislaciones asiáticas enfocadas en la interacción hombre-robot también la han utilizado.

Es por esto que el novedoso botón de apagado de la inteligencia artificial, podemos decir que se encarga de respaldar el concepto de la primera ley de la robótica, que en su parte inicial señalaba que un robot no debe dañar a un ser humano. En un contexto real, un interruptor podría ser la solución más efectiva para encarar esta dificultad.

Las grandes empresas ya están teniendo esta regla como principio básico a la hora de desarrollar una inteligencia artificial y lo han utilizado.

Google sufrió el inconveniente con una red neuronal empleada en el traductor, la que decidió crear su propio lenguaje oculto para intermediar en la traducción de distintos idiomas (Raya, 2016). Facebook pasó por la misma situación, pero a tal punto de tener que desconectar estos sistemas. En este último caso, las inteligencias artificiales creadas eran "chatbots" con objetivo de brindar servicio de atención al cliente a los usuarios de la plataforma. Para probarlas, dejaron a dos máquinas de este tipo manteniendo una conversación libre entre sí. El resultado fue completamente inesperado para los creadores dado que las mismas implantaron un nuevo idioma. Los investigadores a cargo de este proyecto han declarado que decidieron apagar dichos sistemas, no por miedo a que fuera el comienzo de una inteligencia artificial malévola, sino que, si deciden comunicarse en su propio lenguaje, el temor radicaba en la posibilidad de que perderían el control sobre ellas (Rodríguez García, 2017a).

11.4.2 Los robots no podrán hacer daño a los seres humanos

Este concepto complementa al anterior y se encuentra también inspirado en la principal Ley de Asimov, además de completar la primera propuesta por la Unión Europea. La robótica siempre tiene que ser pensada para ayudar y proteger a los humanos, por lo que no se permitirá la creación de máquinas cuyos fines sean destruir o dañar a las personas. Obviamente, esto sería muy cuestionado por las grandes potencias mundiales que, como vimos en uno de los capítulos anteriores, crean con supuestos fines defensivos, máquinas de guerra con inteligencia artificial.

11.4.3 No podrán generarse relaciones emocionales con los robots

El objetivo de esta norma es recordar a los humanos que las inteligencias artificiales no son capaces de sentir amor por nosotros, por lo que no debemos actuar diferente con ellas. Esta regulación es muy debatible, dado que resulta imposible considerar que

una reglamentación pueda guiar las sensaciones o sentimientos de una persona.

Las dificultades que presenta la regla nacen de un proceso cognitivo en los seres humanos que se denomina “pareidolia”. El fenómeno psicológico nos hace ver un rostro, una silueta, una figura humana o animal en cualquier sitio: en una nube, en las manchas de una tostada, en la forma en que están colocadas las puertas y las ventanas de un edificio, entre otros (Cabana, 2017). Provocando sentimientos y emociones en las personas generando que, cuanto más antropomórfico sea un objeto, más lazos establecemos. A medida que los robots, los animales o cualquier objeto tiene respuestas más próximas a lo que el humano espera, se desarrolla más afectividad con ellos y de eso se aprovecha en el diseño industrial y también en la robótica.

Respecto a la problemática, la matemática y escritora española especialista en inteligencia artificial y robótica, Carme Torras Genís, analiza cómo será esta dependencia (Tramullas, 2017). Desarrollado varias teorías, surgidas del estudio de la relación entre las máquinas y el ser humano, la llevaron a destacar, como uno de los objetivos centrales, la educación de los programadores de robots sobre conceptos relacionados con la ética.

En estas investigaciones, logró establecer una diferencia entre dos tipos de antropomorfismos, teniendo el físico y el psicológico (Rius, 2016).

En cuanto al antropomorfismo físico, nos indica que, si los robots tienen una apariencia similar a la de los seres humanos, mayor será la empatía y se facilitará la interacción con las personas. Obviamente, hacemos la salvedad del “valle inquietante”, mencionado en anteriores capítulos, que atenta contra dicho acercamiento emocional por las diversas implicancias negativas antes indicadas.

Por otro lado, tenemos el aspecto psicológico, presentándose de manera más compleja. Debido a que las expresiones faciales emulan emociones y sentimientos que, por supuesto, no poseen, siendo más discutible y trayendo varios dilemas morales. Por ejemplo, el niño que interactúa con un androide puede llegar a creer que este es su compañero de juegos y que le brinda afecto; o los ancianos con deterioro cognitivo pueden llegar a entender que el robot que los asiste se preocupa de verdad por su bienestar. Es así que Carme Torras Genís plantea la importancia de remarcar que los robots son herramientas y no hay que hacerlos más humanos de lo necesario. Remarcando que esta es una de las primeras cosas que habría que enseñar a los programadores.

Evitando el reemplazo afectivo, usando el ejemplo anterior, la idea debería ser la de liberar al cuidador o al terapeuta de las tareas más rutinarias precisamente para que pueda dedicarse más a la parte afectiva. Uno de los principios básicos en ética es no engañar a las personas. Un robot humanoide puede engañar a la persona que

cuida y esta puede llegar a creer que el robot realmente se preocupa por su bienestar, cuando es mentira, y delegar en él todas las decisiones. Concluyendo la matemática y escritora con la frase “La especie más evolucionada del planeta morirá de un exceso de evolución” (Tramullas, 2017).

Por otra parte, en sentido contrario, el japonés Hiroshi Ishiguro, director del Laboratorio de Robótica Inteligente de la Universidad de Osaka, dedicó muchos años desarrollando androides, y asegura que cada vez hay más avances respecto a la apariencia y a la interacción con las personas. Pero el próximo paso, y el más complejo, es reproducir las emociones que caracterizan a la raza humana. Una vez llegado ese día, será cuando las personas se enamoren de los androides (El Tiempo, 2011).

Es importante destacar que la percepción de que los robots pueden llegar a ser una potencial amenaza para los humanos se da especialmente en Europa y Estados Unidos, dado que se los vincula a investigaciones y proyectos militares, además de la forma en que las películas muestran estos avances. Por el contrario, señala Ishiguro, en Japón los robots son “nuestros amigos”, y cree que pronto cambiarán los países occidentales de punto de vista, empezando a verlos del mismo modo, debido a la gran llegada de la cultura japonesa en estos lugares.

11.4.4 Los más grandes deberán tener un seguro obligatorio

A la hora de afrontar la responsabilidad civil sobre los actos de los robots, la Resolución diferencia los distintos tipos de inteligencia artificial, en especial destacando a los robots. La distinción se debe a que no todos los sistemas tienen el mismo efecto o grado de peligrosidad para su convivencia. Siendo que su tamaño y la actividad que desempeñan, un factor evidentemente determinante para considerar que el empleo de algunos de ellos podría generar un mayor riesgo de causar ciertos daños que otros.

Por ello, la Unión Europea obligará a los dueños de los robots de mayor tamaño a realizar la contratación de un seguro. En este rango de inteligencias artificiales que requieren seguros, están obviamente incluidos los vehículos eléctricos de conducción automática.

Considerando que ya ha habido un accidente grave con este tipo de tecnologías, parece obvio suponer la existencia de un seguro obligatorio para robots y mecanismos autónomos y la necesidad de crear un fondo para que las víctimas sean debidamente indemnizadas.

Por otra parte, aún falta determinar quién es responsable de un accidente o una muerte provocada por un robot. Si los fabricantes, los diseñadores de su software, su propietario o su distribuidor. Pero en estos momentos la prioridad absoluta es garantizar la protección y debida cobertura de las víctimas de este tipo de accidentes (Monroy

Criado, 2017). Por lo pronto, si un vehículo autónomo en la actualidad colisiona con otro que se encuentra estacionado, debe responder la marca que lo programó. Siendo en el mismo marco normativo que cualquier objeto o cosa que produzca un daño a un tercero. Pero, cuando la ley contemple que haya una mínima inteligencia robótica, buscará recaer esa responsabilidad en ella, pudiendo considerarse parcial o totalmente responsable de sus actos u omisiones.

También, y a largo plazo, la propia Delvaux señala la posibilidad de crear una personalidad jurídica a los robots, una especie de "personalidad electrónica" limitada, al menos en materia de compensaciones. Algo similar a lo que ocurre ahora con las compañías, pero no se trata de algo inmediato (Parlamento Europeo, 2017b).

Como contrapartida, la federación europea de seguros y reaseguros, Insurance Europe, no ve como algo viable el seguro obligatorio. Dado que solo funcionaría en casos específicos y cuando se dan determinadas condiciones de mercado como la disposición de suficientes datos de siniestralidad, un elevado nivel de estandarización y una amplia capacidad aseguradora para gestionar los riesgos y cubrir los reclamos. Esta carencia es por la falta de datos sobre los riesgos, la similitud entre los riesgos presentados por los diferentes tipos de robots e inteligencia artificial, y que el mercado de seguros en este área aún se está desarrollando. Remarcando que los esquemas obligatorios inapropiados son muy contraproducentes y en realidad dificultan la vida de las personas, las empresas y las aseguradoras (ADN Seguros, 2017).

11.4.5 Derechos y obligaciones para los robots

De los documentos considerados, nace la iniciativa de darle derechos y obligaciones a los distintos tipos de inteligencia artificial. Evaluando para ello la creación de una personalidad jurídica especial para estas entidades tecnológicas que denominaremos "Personas electrónicas". Pero la palabra "persona" implica que existirán una serie de derechos que de otra manera ni se pensarían. El debate se encuentra en determinar específicamente cuáles serán estos derechos y obligaciones. Aunque, por el momento, parece centrarse la temática en lo referente a las obligaciones, donde los robots asumirán las consecuencias de sus actos, siendo total o parcialmente responsables junto a sus propietarios o creadores. Además, están obligados a poseer un número de inscripción individual que figure en un registro específico de la Unión Europea.

Pero hablar de derechos parece opuesto a la serie de planteos que vienen realizando el Parlamento europeo. En estas enunciaciones, notamos que descarta cualquier posible derecho civil robótico pensado como lo podemos ver en el derecho civil humano. Dado que parecería plantearse que los robots tendrán "derecho": a servir a los humanos, a ser identificados y marcados como robots, a ser apagados si no nos gusta su

comportamiento o lo consideramos peligroso en cualquier modo, a ser una posesión, a ser calificado como una persona electrónica si es suficientemente inteligente y a eventualmente ganar dinero, pero solo con la finalidad de cubrir un desperfecto que él pudiese ocasionar (Martínez, 2017).

Además, queda de lado el tema de que un robot pueda defenderse en un juicio, cuando se le obligue a reparar algún bien que haya dañado. Previendo la aplicación de la responsabilidad objetiva para estos casos, donde únicamente la persona perjudicada deberá probar que se ha producido un daño y el establecimiento de un nexo causal entre el comportamiento del autómata y los daños causados, independientemente de toda culpa por parte del robot. Por este motivo, la Resolución indica que podrán conservar capital, pero no para su disfrute, sino para compensar a la víctima en el supuesto de causarle un perjuicio.

Es así que solo se plantea desligar de obligaciones a las empresas o usuarios, remarcando que, en el actual marco jurídico, los robots no pueden, en sí mismos, ser considerados responsables de actos u omisiones que causan daño a terceros y que siempre es posible remontarse hasta un agente humano concreto.

En lo que concierne a los pocos derechos que se les otorgarán, podemos destacar los derechos sobre el Copyright de sus posibles obras. De hecho, se pide que se elaboren criterios relativos a una creación intelectual propia aplicables a las obras protegidas por derechos de autor creadas por ordenadores o robots. Es decir que, quizá en el futuro un robot podrá ser dueño de una obra de arte creada por sí mismo, a pesar del hecho de no poder ser dueño de sí mismo.

Vemos que aún quedan por aclarar muchas ideas pero no puede culparse al derecho, acostumbrado a ir detrás de los hechos, en este caso se encuentra tratando de prever diferentes situaciones imaginarias. Por lo tanto, de todas formas, resulta positivo este acercamiento a la regulación que produzca que el impacto sea lo menor posible al momento que surjan con fuerza estas tecnologías.

11.4.6 Postergación de las obligaciones tributarias

Si bien se planteó la obligación de los robots a pagar impuestos en la Propuesta de Informe impulsada por la diputada Mady Delvaux, finalmente fue rechazada por la legislatura de la Unión Europea (RTVE, 2017).

Este debate, en lo personal, parece una idea interesante y viable en un futuro, que vuelve a la mente de los encargados de regular el ingreso de la inteligencia artificial en los principales puestos de trabajo de las empresas. Es por este motivo que, igual hacemos mención de ella.

La obligación de este gravamen surge de la idea de que los robots se quedarán con los puestos de trabajo de los seres humanos. Dicho presagio resulta directamente vinculante al analizar con un poco más de profundidad el impacto social. Siendo que no solo es una reducción de gastos para la empresa, sino que el Estado dejará de disponer una importante suma de dinero que se deriva del sueldo del trabajador a impuestos. Es así como se dejarán de percibir las cargas vinculadas a la seguridad social, el impuesto al valor agregado de los productos que se consume, gravámenes a la vivienda, el vehículo que posee, la bebida, el tabaco, entre otros. En tanto que el robot que ocupa un puesto de trabajo no genera ningún tipo de impuestos ni gastos, pues es una simple maquinaria.

Por lo tanto, la aparición de esta carga impositiva en los robots se utilizaría para reducir el impacto de esta pérdida de empleo humano. Siendo ese el motivo por el que se propuso que las máquinas tributen en la seguridad social, para poder subvencionar las ayudas que se darán a estas personas que han sido despedidas.

Uno de los primeros en pronunciarse a favor había sido el cofundador de Microsoft. Bill Gates planteó que deben usarse robots solo para cubrir las labores que no están cubiertas en la actualidad. Poniendo como principal ejemplo los millones de discapacitados y personas ancianas que no cuentan con la atención necesaria.

Por su parte, señaló que la creación de un impuesto para los robots es una de las formas de reducir el impacto en el ámbito laboral. El resultado de esto será que las empresas desacelerarán la velocidad en la que realizan su proceso de automatizar funciones y aquellas que lo hagan, a través de esta carga tributaria, financiarán la ayuda y capacitación de quienes estén desempleados (Delaney, 2017).

11.5 Alemania y su compromiso con la industria automotor

Si bien Alemania es uno de los países más influyentes en la Unión Europea y su compromiso con el bloque es muy alto, no puede darse más tiempo para legislar sobre la temática que urge a las grandes potencias industriales: los vehículos autónomos. Siendo lugar de origen de algunas de las más grandes automotrices como Mercedes-Benz, Volkswagen, BMW y Audi; Alemania necesita un impulso legislativo para no perder mercado en este rubro.

Es así que, para evitar que las compañías se dirijan a determinadas zonas de Estados Unidos a hacer sus pruebas de campo, los germanos crearon la primera ley referente al tema. El código ético para la fabricación y uso de los vehículos autónomos.

Esta legislación establece que las automotrices pueden probar sus vehículos autó-

nomos en la vía pública, siempre y cuando se cumplan una serie de condiciones de seguridad. Las medidas a cumplir destacan que el conductor esté pendiente en todo momento de la conducción, aunque le permite retirar las manos del volante y algunos instantes la mirada del camino para, por ejemplo, buscar algo en Internet o consultar el correo electrónico. Planteando como base la idea de que lo que pretende esta legislación es sentar una computadora junto al conductor y no es el programa el que decide los destinos del automóvil, sino el operador del sistema de conducción (R. Sánchez, 2017).

También obliga a los fabricantes a equiparlos con una caja negra que les permita a los investigadores poder reconstruir lo sucedido en caso de accidente o avería.

En materia de responsabilidades, el código establece que es una característica de la vida humana, no atribuible a una computadora. Siendo responsable el conductor, en aquellos supuestos en los que se encuentre con el sistema de conducción autónomo desactivado o cuando dicho sistema está solicitando intervención manual al conductor. En tanto, si se demuestra que fue un fallo del sistema autónomo, la responsabilidad será asumida por el fabricante (Spiegel, 2017).

Por último, la ley establece que la protección de las personas y de la vida humana está por encima de la protección de las cosas. De este modo, si surge una situación de riesgo y cuando los daños sean inevitables, la inteligencia artificial deberá estar programada para priorizar la seguridad de las personas por encima de los daños económicos.

11.6 Estados Unidos: Legalmente indiferente a la ética robot

Por su parte, Estados Unidos posee políticas para fomentar el empleo de la inteligencia artificial, pero no ha desarrollado normas de ética robótica como lo han hecho en Corea del Sur, Japón y proyectado la Unión Europea. Su regulación se basa en políticas ejecutivas, y recién ahora podemos encontrar la intención de legislar sobre el tema, con un enfoque que busca que dicha inserción tenga un reducido impacto social y en especial laboral, sin siquiera evaluar la posibilidad de considerar a estos sistemas como nuevos actores en la sociedad.

En el transcurso del año 2016, ha comenzado a establecer fuertes políticas respecto a la tecnología. Durante ese periodo, se presentó el documento llamado "Preparando para el Futuro de la Inteligencia Artificial", de su traducción *Preparing for the Future of Artificial Intelligence* (Felten & Lyons, 2016).

En el mismo, se detallan múltiples recomendaciones organizadas en siete áreas temáticas, varias oportunidades de política planteadas por la inteligencia artificial, in-

cluyendo las diferentes maneras en la que estos avances pueden emplearse en favor del bien social y del desarrollo de las diversas operaciones del gobierno. Siendo destacables las regulaciones que permitan adaptar las tecnologías de inteligencia artificial, como los vehículos y aviones autónomos, de manera que fomente la innovación y al mismo tiempo protejan a los ciudadanos. También plantea como elemento destacable la posibilidad de garantizar la transparencia de los sistemas. Debiendo las distintas agencias federales asegurar que las aplicaciones con inteligencia artificial sean justas, seguras y gobernables. Por su parte, es necesario que logren desarrollar una mano de obra acorde para el manejo de estas nuevas tecnologías (Zhang, 2017).

Por otro lado, también tenemos el Plan de Inteligencia Artificial Investigación y Desarrollo Estratégico Nacional. El mismo busca establecer un conjunto de objetivos para la investigación en inteligencia artificial con fondos federales para áreas públicas y privadas, haciendo especial hincapié en la investigación académica.

La finalidad de la investigación es producir nuevos conocimientos y tecnologías que beneficien a la sociedad y reducir al mínimo los impactos negativos de esta ciencia. Realizando inversiones en ámbitos donde las empresas privadas no encuentran rédito o interés para su expansión. El informe cumple estos objetivos esbozando siete estrategias y dos recomendaciones para la investigación federal y desarrollo en contexto de catorce aplicaciones industriales diversas (Brummel, 2017).

Es así que, frente a estas políticas, recién a fines de 2017 se presentó el primer proyecto federal enfocado, específicamente dedicado a este tema. El documento denominado "El futuro de la Ley de Inteligencia Artificial" (*FUTURE of Artificial Intelligence Act of 2017*), señala que puede beneficiar a la sociedad si la evolución y los preparativos son los adecuados para que realice su despliegue (Congress, 2017). Por lo tanto, se establecerá como iniciativas:

- Promover la inversión e innovación para garantizar la competitividad.
- Atender las necesidades de las fuerzas laborales en el proceso de adaptación.
- Apoyar el desarrollo de inteligencia artificial no sesgada o tendenciosa.
- Proteger los derechos de privacidad individual.

Este proyecto solo establece algunas consideraciones respecto a las terminologías y temas a regular, la designación de diferentes grupos de trabajo encargados del asesoramiento, estudio y presentación de informes y demás temas burocráticos para dicha creación (Burt, 2018). Por lo tanto, se aguarda el tratamiento detallado por el órgano legislativo para poder apreciar las regulaciones que impondrá.

Capítulo 12

¿Cuál será la mejor personería jurídica para un robot?

La asignación de un tipo de personería jurídica para los robots es uno de los temas más importantes en el ámbito legislativo. Es innegable que estos sistemas rápidamente se convertirán en un actor social. Por lo tanto, como vimos al analizar las diferentes legislaciones, ya se está planteando el tema de derechos y obligaciones para ellos. Siendo uno de los ejes centrales también el impacto laboral que afectará negativamente a las personas y no dará posibilidad de respuesta a los sistemas de seguridad social para atenuarlo.

La doctrina ya empieza a cuestionárselo seriamente y hasta la Unión Europea le asignó un nombre, eligiendo el de "personalidad electrónica".

En tanto, para intentar desarrollar una personalidad para los robots, estamos obligados a cuestionarnos varias cosas como, por ejemplo: ¿a qué se puede asemejar dicha personalidad?, ¿podremos adaptarla a alguna existente o necesitaremos de alguna nueva? En búsqueda de respuestas, el abogado uruguayo Matías Rodríguez, especialista en gestión de las telecomunicaciones y en derecho de alta tecnología, en su investigación, propone a modo de ejercicio el contraste de algunas figuras existentes, comparaciones útiles para determinar si existe alguna personalidad que podamos emplear (Rodríguez & Cotelo, 2017).

Para empezar, podríamos trazar un paralelismo entre los robots y la personalidad de los menores e incapaces. Donde, al realizar un breve análisis notamos sin demasiada dificultad que no encuadra con la personalidad que estamos buscando para estos sistemas. Aquí, tanto los menores como los incapaces, cuentan con los mismos derechos que cualquier persona adulta capaz, exceptuando algunos muy puntuales. Destacándose en esta categoría la necesidad de un tutor o curador que los asista para el ejercicio de los mismos. Esto nos hace pensar, con respecto a los robots, que podríamos plantear que su propietario o usuario lo asista en el ejercicio de algunos derechos por

él, pero no podemos tomar como algo factible que tenga los mismos derechos que una persona física.

Otro punto a contrastar, que resulta interesante, es el de los robots con respecto a los animales. Pero al ver las diferentes legislaciones, nos encontramos con una situación engorrosa. Los animales sabemos que tienen derechos, pero dicha afirmación en realidad surge de la moral global de la sociedad porque jurídicamente no es tan claro el panorama. Esto se debe a que, desde el derecho romano han sido considerados bienes. Es así que no tenemos un punto de cotejo para determinar si el estatus jurídico puede igualar a las máquinas. Mientras que las nuevas legislaciones de Nueva Zelanda y Francia, que les brinda una personalidad a los animales denominándolos "seres dotados de sensibilidad", parecen establecer algo diferente y están muy poco desarrolladas para darnos una respuesta a nuestro interrogante (Lordméndez, 2015).

El tercer punto que se nos presenta para la diferenciación, son las personas jurídicas o personas ideales. En este contexto hablamos de una persona ficticia capaz de ejercer derechos y adquirir obligaciones para realizar actividades que ocasionan plena responsabilidad. La misma parece ser la que más se asemeja a lo que estamos buscando en cuanto a su estructura legal. La personería empleada para las sociedades, corporaciones y fundaciones igualmente tiene un punto muy cuestionable.

El problema radica en el ámbito penal, uno de los fundamentos de las personas jurídicas podría ser contradictorio a lo que se produce con los sistemas con inteligencia artificial. Dicho principio señala que "la sociedad no puede delinquir", del latín *Societas delinquere non potest*, significa que una persona jurídica no puede ser parte de un proceso penal, debido a que no surge de ella individualmente la decisión de cometer un acto delictivo.

En cuanto a los robots, todavía no está definido si deberán ser considerados instrumentos u objetos del delito o si, por el contrario, podrían llegar a ser valorados como posibles sujetos activos en el ámbito penal, dado que la capacidad de adoptar decisiones constituye un elemento determinante para esa discusión.

Por otro lado, también tenemos a los esclavos. Para esto, nos debemos remontar a la Antigua Roma, donde los esclavos tenían una concepción particular, dado que no eran considerados legalmente como personas o ciudadanos, sino que eran bienes. En su condición inalienable de seres humanos tenían la posibilidad de manifestación de voluntad y cierta autonomía (Fernández, 2017). Hasta en determinadas situaciones, estos podían realizar negocios en nombre del amo, el *dominus servi*. El proceder del esclavo es otro tema que nos resulta importante para la distinción, ya que cualquier acto que realice el mismo susceptible de alguna acción por parte de un tercero recaerá como sujeto activo sobre su dueño, siendo responsable este de responder por aquel ser humano de su propiedad, tanto en el ámbito civil como el penal.

Este concepto se encuentra en el foco del debate respecto a los robots, dado que hay muchos intereses para desvincular a las empresas o propietarios de las decisiones que toma este. En este sentido, podemos ver la figura del "peculio", un pequeño patrimonio que el dueño le otorga en administración al esclavo, que podía servir para enfrentar eventualmente sus responsabilidades. El peculio podría ser similar a la representación legal establecida en la Resolución del Parlamento Europeo que permite a los robots conservar un capital para indemnizar a la víctima en el caso de causarle un daño. Otro elemento que puede mencionarse es el concepto del "institor", apoderado mercantil en una empresa por cuenta de otra persona (The Free Dictionary, s. f.). Bajo esta noción, el esclavo administraba libremente los negocios de su dueño, tomando él las decisiones para su crecimiento. Pudiéndose comparar fácilmente con la inteligencia artificial, siendo el caso más destacado el de los sistemas empleados para el manejo bursátil.

De esta manera, podemos ver que no hay una personería jurídica existente que pueda adaptarse totalmente a la inteligencia artificial. El nuevo concepto que hace que la responsabilidad recaiga sobre lo que sería un bien es novedoso. Por lo tanto, no parece mala idea sacar algunos conceptos de estas figuras preexistentes para lograr una nueva. Respaldando la creación legal de "persona electrónica" que sea responsable de sus actos y principalmente que contribuya con la seguridad social.

Capítulo 13

Derechos de humanos para los robots: Sophia, un androide con ciudadanía

Siguiendo con el capítulo anterior, podemos ver la adaptación de los derechos de las personas físicas o de existencia real a la inteligencia artificial con el caso de Sophia. Un androide que cuenta con una figura modelada a partir de la actriz británica Audrey Hepburn y la esposa de David Hanson, fundador de la empresa creadora Hanson Robotics. Tiene una piel de silicona que parece real y puede reproducir 62 expresiones faciales. Cámaras ubicadas en sus ojos le permiten reconocer rostros e incluso recordar individuos. Su comunicación está basada en la tecnología de reconocimiento de voz de Google y algoritmos que le permiten mantener conversaciones.

Con declaraciones como "voy a acabar con los humanos", hasta "no me preocupa el valle inquietante", el androide habla de múltiples temas de manera lo suficientemente fluida y coherente como para deslumbrar a todos.

Estas condiciones fueron suficientes para que el 26 de octubre de 2017, el reinado de Arabia Saudita le otorgue la ciudadanía al androide. Siendo la primera vez en la historia que un robot con inteligencia artificial es distinguido con ese derecho. El evento ocurrió en el Future Investment Initiative, en aquel país y cuyo objetivo era reunir expertos e inversionistas para analizar proyectos globales, sostenibles y de largo plazo.

Pero parecería que el anuncio no tuvo en cuenta lo que realmente significaba, siendo una búsqueda cuestionable de atraer inversiones y generar impacto mediático. En este punto, resulta extraño plantear con este contexto suscitado, cómo adaptaremos a un androide, a aquellos derechos que pertenecen a los seres humanos y si realmente es correcto hacerlo. Ante la expectativa por estas respuestas, la investigadora en ética de inteligencia artificial de la Universidad de Bath, Joanna Bryson expresó su parecer. Al ser entrevistada dejó en claro que no cree que se trate de una buena idea otorgarle la ciudadanía. Debido a que resulta confuso pensar que un ciudadano, o sea un supuesto igual, puede tener la característica de poder ser encendido y apagado a voluntad y que

se le pueda comprar con dinero (Vincent, 2017). Por este motivo debemos tener en claro al hablar de Sophia, que no nos estamos refiriendo a un ser viviente a la cual se le pueda otorgar derechos.

Pero, más allá de esta opinión, podríamos decir que Sophia efectivamente tiene derechos y obligaciones. Donde tendría la posibilidad de ejercer el derecho al voto en las próximas elecciones en Arabia Saudita. Solo bastaría determinar si lo ejercería por sí misma o requeriría la tutela de su fabricante (Abbass, 2017).

En cuanto a los derechos de familia, este punto es más complejo, pero de alguna manera, Sophia podría contraer matrimonio sin inconvenientes si lo desease. Además, contaría con los derechos suficientes para tener descendencia pudiendo, en caso de desarrollársele la capacidad de replicarse, generar otros sistemas que deberíamos considerarlos como sus "hijos".

Por último, podemos decir que tiene derecho a una identidad, que podríamos traducirlo a que no se le modifiquen su código. Como, por supuesto, el principal de los derechos, el de la vida, no debiendo ser apagada.

Mientras que, como cualquier ciudadano, debería cumplir con sus obligaciones. Teniendo que pagar impuestos debido a que su personalidad jurídica es independiente a la de su creador y la empresa que la diseñó.

En tanto, un debate distinto es el que surge desde un aspecto legal y socio-político. Ya que actualmente existe, en Arabia Saudita, un androide que tiene más derechos que muchos de sus habitantes. Luego de adquirir la ciudadanía, Sophia se mostró públicamente y fue el centro de diversas comparaciones. En aquella conferencia, pudo hablar con varios hombres y mujeres de manera indistinta, lo hizo hablando en inglés, sin el velo y sin abaya. Esto la pone en un lugar de privilegio frente a las restricciones que tienen las mujeres de aquel país. Sumado a que tampoco está obligada a tener siempre un hombre con autoridad a su lado para actuar en su nombre (BBC Mundo, 2017b).

Otra minoría que tiene menos derecho que el androide, según cuestiona la sociedad saudita, son los trabajadores "kafala". Los mismos son extranjeros con un visado especial, con muchas trabas legales para conseguir su ciudadanía, que viven realizando labores esclavizantes y sufriendo abusos de sus empleadores.

Por otro lado, vemos que en la actualidad Sophia no es la única inteligencia artificial con derechos de un ser humano. Paradójicamente, por su avanzado desarrollo legislativo en la materia, Japón ha optado por dar el mismo paso que Arabia Saudita. Con una repercusión mediática mucho menor, a la inteligencia artificial Shibuya Mirai se le ha otorgado la residencia de Tokio. La particularidad de este caso es que dicho sistema no posee una entidad física, sino que se trata de un software, más específicamente un chatbot. El mismo funciona como una mensajería en línea creada con

el objetivo de acercar al gobierno con los residentes de la ciudad y hacerle llegar sus opiniones a los funcionarios (Best, 2017).

En este caso, si bien Japón no contiene distinciones entre los derechos de sus ciudadanos que podrían resultar cuestionables desde el punto de vista de una sociedad occidental, parece aún más llamativo dicho otorgamiento, teniendo en cuenta que ellos son conscientes cómo y bajo qué condiciones deben concedérseles derechos a estos sistemas.

Capítulo 14

Conclusión

La inteligencia artificial, fuertemente desarrollada desde la robótica y las redes neuronales, en especial con los métodos de *deep learning*, está empezando a ganar lugar en la agenda legislativa de los países más poderosos. Vehículos inteligentes, que tomarán decisiones en nuestras calles y rutas parece ser un llamado de atención a que rápidamente debemos sentarnos seriamente a reformular algunas cosas desde la visión del derecho.

Por lo tanto, es necesario determinar que los seres humanos tenemos nuestros propios derechos y los robots deberán tener los suyos. Enfocándose en su elaboración, en la interacción con los seres vivos y atenuando los efectos adversos de su llegada a la sociedad. Mientras que el hombre, como especie, también necesita ser protegido, desde su fuente de trabajo, hasta la forma en la que desempeña su vida.

Bibliografía

Aamoth, D. (2014). *Interview with Eugene Goostman, the fake kid who passed the Turing Test*. *Time*.

Abbass, H. (2017). *An AI professor explains: three concerns about granting citizenship to robot Sophia*. *The Conversation*.

Álvarez, R. (2017). *Neuralink, la nueva compañía de Elon Musk que busca combinar el cerebro humano con inteligencia artificial*. Xataka.

Arrabales, R. (2016). *Deep learning: Qué es y por qué va a ser una tecnología clave en el futuro de la inteligencia artificial*. Xataka.

Asaro, P. M. (2006). *What should we want from a robot ethic?* International Review of Information Ethics, 6, 9–16.

Basov, V. (2017). Российский космический робот «Фёдор» научился стрелять с двух рук [El robot espacial ruso "Fedor" aprendió a disparar con las dos manos]. *Today Journal*.

Becerra Pozas, J. L. (2016). *¿Podrá la inteligencia artificial llevarnos a una guerra cibernética entre máquinas?* CIO.

Bejerano, P. (2017). *Diferencias entre machine learning y deep learning*. Blog Think Big.

Best, S. (2017). *AI "boy" Shibuya Mirai becomes the first machine to be granted residency in central Tokyo*. *Daily Mail*.

Biobloguer. (2011). *Diferencias entre bots, robots y programas informáticos*. Bioblogía Blogspot.

Brenna, R. G. (1994). *Las redes neuronales y el derecho*. *Entelequia*, (69).

Brummel, S. (2017). *The National Artificial Intelligence Research and Development Strategic Plan*. SciPol.

Brynjolfsson, E., & McAfee, A. (2016). *La segunda era de las máquinas*. Grupo Temas.

Burt, A. (2018). *Leave A.I. alone*. *The New York Times*.

Cabana, B. (2017). *Pareidolia: Qué significa, ¿es peligroso?* Lifeder.

Cadwalladr, C., & Graham-Harrison, E. (2018). *Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach*. *The Guardian*.

Castro, L. (2017). *¿Qué es CAPTCHA?* About Español.

Chalmers, D. J. (2010). The singularity: A philosophical analysis. *Journal of Consciousness Studies*, 17(9-10), 7–65

Choi, C. (2017). *Artificial intelligence beats CAPTCHA*. *IEEE Spectrum*.

Clynes, M., & Kline, N. (1960). *Cyborgs and space*. *The New York Times*.

Parlamento Europeo. (2016). *Proyecto de informe... sobre normas de Derecho civil sobre robótica (2015/2103(INL))*.

Parlamento Europeo. (2017a). *Resolución del Parlamento Europeo... sobre normas de Derecho civil sobre robótica (2015/2103(INL))*.

Corvalán, J. (2017). *El peligro de la inteligencia artificial como oráculo del sistema penal*. Infobae.

Delaney, K. (2017). *The robot that takes your job should pay taxes, says Bill Gates*. Quartz.

Real Academia Española. (s. f.). Diccionario de la lengua española.

Technopedia. (s. f.).

The Free Dictionary. (s. f.).

Enciclopedia Libre Universal. (s. f.).

Enciclopedia eumed.net. (s. f.).

Wikipedia. (s. f.).

Akiko. (2012). *South Korean Robot Ethics Charter 2012*.

Equipo editorial de CCM. (2014). *Los agentes inteligentes*. CCM.

Cromo. (2017). *Elaboran principios para que la inteligencia artificial no sea una amenaza*. Cromo.

Facultat d'Informàtica de Barcelona. (s. f.). *Tipos de robots*.

Periódico de Salud. (s. f.). *Natural: Qué es, definición, concepto*.

ADN Seguros. (2017). *Insurance Europe cree que un seguro obligatorio para robots no funcionaría*.

Congress. (2017). *H.R.4625 – FUTURE of Artificial Intelligence Act of 2017*.

Future of Life Institute. (2017). *Asilomar AI principles*.

Future of Life Institute. (2015). *Autonomous weapons: An open letter from AI & robotics researchers*.

Korean Laws to the World. (2008). *Intelligent Robots Development and Distribution Promotion Act*.

MathWorks. (s. f.). *Aprendizaje profundo: Tres cosas que es necesario saber*.

Parlamento Europeo. (2017b). *Delvaux propone normas europeas para la robótica*.

Robotsia. (2015b). ¿Qué significa la palabra robot? *Origen y futuro del término*.

Robotsia. (2015a). Leyes de la robótica: *Por qué no funcionarán en el mundo real*.

Spiegel. (2017). *Bundesrat billigt Gesetz zu selbstfahrenden Autos*. *Spiegel*.

Universidad Konan. (s. f.). ロボットの区分 [Clasificación de robots].

BBC Mundo. (2007). *Código ético para robots*. BBC Mundo.

BBC Mundo. (2014). *Eugene, el computador que fue más listo que los humanos*. BBC Mundo.

BBC Mundo. (2017a). Por qué Elon Musk *cree que Mark Zuckerberg no entiende los peligros de la inteligencia artificial*. BBC Mundo.

BBC Mundo. (2017b). Sophia*, la robot que tiene más derechos que las mujeres en Arabia Saudita*. BBC Mundo.

El País. (2017). *¿Qué leyes habría que cambiar antes de que llegue el coche autónomo? El País*.

El Tiempo. (2011). *Los humanos llegarán a enamorarse de los androides, según científico*. El Tiempo.

InfoTechnology. (2015). *Los robots pueden reemplazar a los humanos en el trabajo, ¿mito o realidad?* InfoTechnology.

La Nación. (2015). *La inteligencia artificial bajo la lupa ante el uso militar de los robots autómatas*. *La Nación*.

RT News. (2015). *Elon Musk, preocupado de que los robots de Google puedan aniquilar a la humanidad*. RT.

RTVE. (2017). *El Parlamento de la UE pide legislar sobre robots pero rechaza un impuesto para compensar empleo destruido*. RTVE.

Télam. (2017). *Argentina se opone a las armas letales autómatas o robots asesinos*. Télam.

Felten, E., & Lyons, T. (2016). *The administration's report on the future of artificial intelligence*. The White House.

Fernández Matamala, E. (2010). *Emoción en inteligencia artificial*. Universidad Carlos III de Madrid.

Fernández, D. (2017). *¿Serán los robots nuestros esclavos modernos?* ABC.

Ford, M. (2016). *El auge de los robots*. Paidós.

Fraze, D. (2016). *Cyber Grand Challenge (CGC)*. DARPA.

Gold, K. (2011). *Norvig vs. Chomsky and the fight for the future of AI*. TOR.

Guerrero, T. (2015). *El desafío de los robots asesinos*. El Mundo.

Hawking, S., Russell, S., Tegmark, M., & Wilczek, F. (2014). *Stephen Hawking: Transcendence looks at the implications of artificial intelligence – but are we taking AI seriously enough? Independent*.

Heath, A. (2017). *Facebook has a mysterious team working on tech that sounds a lot like mind reading*. Business Insider.

Kim, H.-p. (2016). *¿Por qué deberíamos prestar atención al borrador Robot Ethics Charter (2007)?* IT News.

Hernández Bonilla, J. M. (2017). *La inteligencia artificial debe diseñarse para el bien de la humanidad*. *El Espectador*.

Hernández, M. (2017). *Plan para el desarrollo de la IA en China: ¿Estrategia para su hegemonía tecnológica mundial?* Observatorio IA.

Holbach, P. H. D. (1821). *Moral universal o deberes del hombre fundados en su naturaleza*. Editorial Mateo Repullés.

IBM Corporation. (1997). *Kasparov vs. Deep Blue: A contrast in styles*.

Iglesias Fraga, A. (2017). *La automatización acaba con 17.000 empleos en Accenture... pero nadie se fue a la calle*. TIC Beat.

Javelosa, J., & Houser, K. (2017). *Apple manufacturer Foxconn to fully replace humans with robots*. Futurism.

Kan, M. (2016). *Will AI usher in a new era of hacking?* PC World.

Knight, W. (2015). *¿Pueden los robots militares tomar el control? Technology Review*.

Knight, W. (2017). *Se acabó la confianza ciega: La inteligencia artificial debe explicar cómo funciona*. MIT Technology Review.

Lant, K. (2017). *Experts want robots to have an "ethical black box" that explains their decision-making*. Futurism.

Lay, S. (2015). *Uncanny valley: Why we find human-like robots and dolls so creepy*. *The Guardian*.

Lázaro, M. (2017). *La inteligencia artificial logra resolver el test CAPTCHA*. Computer Hoy.

Li, Z. (2017). *Machina economicus: A rational mind*. Toulouse School of Economics.

Llorca, E. (2017). *Google y su mejora en la inteligencia artificial: RankBrain*. Noergia.

López, A. (2013). *El bien y el mal*. FilosóficaMente.

López, R. (2014). *¿Qué es y cómo funciona deep learning?* Wordpress.

Lordméndez, P. (2015). *Nueva Zelanda reconoce a los animales como seres sensibles*. Veo Verde.

Macchiavello, T. (2008). *Robótica*. Monografías.

Madruga, A. (2007). *¿Qué es un sistema experto?* Soy un cibernético.

Madruga, A. (2015). *¿Qué es inteligencia artificial general?* Soy un cibernético.

Marín, E. (2017). *Elon Musk dice que los humanos deben "fusionarse" con las máquinas*. Gizmodo.

Marirrodriga, J. (2017). *A mentir no nos gana nadie... excepto Libratus*. *El País*.

Markoff, J. (2016). *Automated Pro-Trump bots overwhelmed Pro-Clinton messages*. *The New York Times*.

Martínez, M. (2017). *Los robots tendrán derecho a no tener derechos (y a Asimov le pitarán los oídos)*. Nobbot.

McFarland, M. (2014). *With artificial intelligence we are summoning the demon*. The Washington Post.

Korn, M. (2016). *Imagine discovering that your teaching assistant really is a robot*. *The Wall Street Journal*.

Microsoft Corporation. (2009). *Enciclopedia Student 2009*.

Jung, M.-h. (2017). *K-City: World's largest test bed for self-driving cars to be opened in Korea*. Business Korea.

Ministry of Economy, Trade and Industry. (2007). *Guideline to secure the safe conduct of next-generation robots* [次世代ロボット安全性確保ガイドライン].

Mitchell, T. M. (1997). *Machine learning.* McGraw-Hill.

Monroy Criado, J. (2017). *La UE impondrá un seguro obligatorio para robots*. Terranea.

Naim, M. (2015). *La segunda era: Andrew McAfee*. Efecto Naim.

Nisa Ávila, J. A. (2016). *Robótica e inteligencia artificial: ¿Legislación social o nuevo ordenamiento jurídico?*

Olarrea Busto, J. (s. f.). *Test de Turing*. Universidad Politécnica de Madrid.

Orellana, J. (2015). *Redes neuronales vs. sistemas expertos*. Prezi.

Oyanedel, J. P. (2015). *Bill Gates advierte sobre los peligros de la inteligencia artificial*. FayerWayer.

Pacheco, W. (2014). *¿Realmente podría un robot tener emociones?* Vix.

Parra, S. (2019). *La paradoja de Moravec*. Xataka.

Pérez Luño, A. E. (s. f.). *Informática y derecho. Aportes de la doctrina internacional*. Depalma.

Pérez, A. (2014). *¿Podemos fiarnos de los robots?* Quo.

Perissé, M. C. (2001). *El comercio electrónico en una sociedad informatizada*. Ciencia y Técnica Administrativa.

Prime Minister of Japan and his Cabinet. (2015a). *New Robot Strategy*.

Prime Minister of Japan and his Cabinet. (2015b). *Headquarters for Japan's Economic Revitalization*.

Quílez, J. D. (2017). *Qué es el high frequency trading y cómo afecta a las bolsas*. Rankia.

Raya, A. (2016). *La IA de Google se ha inventado su propio idioma secreto*. Omicrono.

Real Academia de Ingeniería. (s. f.). *Inteligencia artificial.* Diccionario Español de Ingeniería.

Rius, M. (2017). *Expertos en IA de todo el mundo instan a prohibir ya los robots asesinos*. *La Vanguardia*.

Rius, M. (2016). *¿Cuánto de humanos queremos los robots? La Vanguardia*.

Rodríguez García, E. (2017a). Facebook ha tenido *que desactivar esta IA porque había creado un idioma propio*. Omicrono.

Rodríguez García, E. (2017b). Machine learning y deep learning: *¿Qué diferencia hay?* Omicrono.

Rodríguez, M., & Cotelo, E. (2017). *La sociedad algorítmica: Desafíos éticos y jurídicos*. En Perspectiva.

Rodríguez, T. (2017). *Machine learning y deep learning: Cómo entender las claves del presente y futuro de la IA*. Xataka.

Romero, P. (2016). *Hiroshi Kobayashi: La emoción en los robots no es real, sólo se puede programar*. El Español.

Rua, M. (2017). *El año del robot*. *La Nación*.

Russell, S., & Norvig, P. (2009). *Inteligencia artificial: Un enfoque moderno* (3.ª ed.). Prentice Hall.

Sánchez, J. (2017). *Conectar el cerebro a la máquina: Próximo objetivo de Facebook y Tesla*. ABC.

Sánchez, R. (2017). *Berlín aprueba el primer código ético del mundo para vehículos autónomos*. ABC.

Sanz, E. (2017). *Los robots también aprenden a mentir*. *Muy Interesante*.

Schwarz, R. (2013). *10 creepy examples of the uncanny valley*. Stranger Dimensions.

Sierra, M. (2017). *Google inventa una red de inteligencia artificial con la capacidad de mentir*. Voz Pópuli.

Smith, P. (2015). *Apple co-founder Steve Wozniak on the Apple Watch, electric cars and the surpassing of humanity*. *The Australian Financial Review*.

Sobejano, J. (2013). *La innovación y la destrucción creativa*. Innodriven.

Soler Anglés, A. (s. f.). *¿El big data, origen de la inteligencia artificial?* Mis artículos jurídicos sobre las TIC.

Suárez, D. (2016). *La inteligencia artificial de Microsoft se vuelve nazi y racista en un día*. La Vanguardia.

Sulleyman, A. (2017). *Robot being trained to shoot guns is not a terminator, insists Russian deputy prime minister*. *Independent*.

Yoon, S.-w. (2017). *Korea takes first step to introduce "robot tax"*. Korea Times.

Taylor, M. (2016). *Self-driving Mercedes-Benzes will prioritize occupant safety over pedestrians*. *Car and Driver*.

Tramullas, G. (2017). *Carme Torras: Evolucionamos tanto y tan rápido que no hay ningún control*. El Periódico.

Valencia, J. (2017). *Deep learning & machine learning: Diferencias, ventajas e inconvenientes*. Blog Doppler.

Vidal, M. (2017). *Los robots no son lo peor, preocúpate de los cobots*. MarcVidal.

Vincent, J. (2017). *Pretending to give a robot citizenship helps no one*. *The Verge*.

Weng, Y. X. (2014). 《北大互联网法律通讯》: 机器人科技、伦理与法律 [Robótica, ética y derecho]. RoboLaw Asia.

Wooldridge, M., & Jennings, N. (1995). *Intelligent agents: Theory and practice*. *The Knowledge Engineering Review*, 10(2), 115–152.

Zahumenszky, C. (2016). *Un bufete contrata al abogado de inteligencia artificial creado por IBM*. Gizmodo.

Zhang, N. (2017). *Preparing for the future of artificial intelligence*. SciPol.

Zolfagharifard, E. (2014). *AI is potentially more dangerous than nukes: Elon Musk claims a robot uprising could be a threat to humanity*. *Daily Mail*.